Inform, educate, entertain... and recommend?

Exploring the use and ethics of recommendation systems in public service media

November 2022

Contents

This report explores the ethics of recommendation systems as used in public service media organisations. These independent organisations have a mission to inform, educate and entertain the public, and are often funded by and accountable to the public

Executive summary

'Where should I go for dinner? What should I read, watch or listen to next? What should I buy?' To answer these questions, we might go with our gut and trust our intuition. We could ask our friends and family, or turn to expert reviews. Recommendations large and small can come from a variety of sources in our daily lives, but in the last decade there has been a critical change in where they come from and how they're used.

Recommendations are now a pervasive feature of the digital products we use. We are increasingly living a world of **recommendation systems**, a type of software designed to sift through vast quantities of data to guide users towards a narrower selection of material, according to a set of criteria chosen by their developers. Examples of recommendation systems include Netflix's 'Watch next' and Amazon's 'Other users also purchased'; TikTok's recommendation system drives its main content feed.

But what is the risk of a recommendation? As recommendations become more automated and data-driven, the trade-offs in their design and use are becoming more important to understand and evaluate.

Background

This report explores the ethics of recommendation systems as used in public service media organisations. These independent organisations have a mission to inform, educate and entertain the public, and are often funded by and accountable to the public.

In media organisations, producers, editors and journalists have always made implicit and explicit decisions about what to give prominence to, both in terms of what stories to tell and what programmes to commission, but also in how those stories are presented. Deciding what makes the front page, what gets the primetime slot, what makes top billing on the evening news – these are all acts of recommendation. While private media organisations like Netflix primarily use these systems to drive user engagement with their content, public service media organisations, like the British Broadcasting Corporation (BBC) in the UK, operate with a different set of principles and values.

This report also explores how public service media organisations are addressing the challenge of designing and implementing recommendation systems within the parameters of their mission, and identifies areas for further research into how they can accomplish this goal.

While there is an extensive literature exploring public service values and a separate literature around the ethics and operational challenges of designing and implementing recommendation systems, there are still many gaps in the literature around how public service media organisations are designing and implementing these systems. Addressing these gaps can help ensure that public service media organisations are better able to design these systems. With this in mind, this project has explored the following questions:

- What are the values that public service media organisations adhere to? How do these differ from the goals that private-sector organisations are incentivised to pursue?

- In what contexts do public service media use recommendation systems?

- What value can recommendation systems add for public service media and how do they square with public service values?

- What are the ethical risks that recommendation systems might raise in those contexts? And what challenges should teams consider?

- What are the mitigations that public service media can implement in the design, development, and implementation of these systems?

In answering these questions, we focused on European public service media organisations and in particular on the British Broadcasting Corporation (BBC) in the UK, who are project partners on this research.

The BBC is the world's largest public service media organisation and has been at the forefront of public service broadcasters exploring the use of recommendation systems. As the BBC has historically set precedents that other public service media have followed, it is valuable to understand its work in depth in order to draw wider lessons for the field.

In this report, we explore an in-depth snapshot of the BBC's development and use of several recommendation systems from summer and autumn 2021, alongside an examination of the work of several other European public service media organisations. We place these examples in the broader context of debates around 21st century public service media and use them to explore the motivations, risks and evaluation of the use of recommendation systems by public service media and their use more broadly.

The evidence for this report stems from interviews with 11 current staff from editorial, product and engineering teams involved in recommendation systems at the BBC, along with interviews with representatives of six other European public service broadcasters that use recommendation systems. This report also draws on a review of the existing literature on public service media recommendation systems and on interviews with experts from academia, civil society and government.

Findings

Across these different public service media organisations, our research has found **five key findings**:

1. **The contextual role of public service media organisations is a major driver for their increasing use of recommendation systems.** The last few decades have seen public service media organisations lose market share of news and entertainment to private providers, putting pressure on public service media organisations to use recommendation systems to stay competitive.

2. **The values of public service media organisations create different objectives and practices to those in the private sector.** While private-sector media organisations are primarily driven to maximise shareholder revenue and market share, with some consideration of social values, public service media organisations are legally mandated to operate with a particular set of public interest values at their core, including universality, independence, excellence, diversity, accountability and innovation.

3. **These value differences translate into different objectives for the use of recommendation systems.** While private firms seek to maximise metrics like user engagement, 'time on product' and

This report makes nine recommendations for future research, experimentation and collaboration between public service media organisations, academics, funders and regulators

subscriber retention in the use of their recommendation systems, public service media organisations seek related but different objectives. For example, rather than maximising engagement with recommendation systems, our research found public service media providers want to broaden their reach to a more diverse set of audiences. Rather than maximising time on product, public service media organisations are more concerned with ensuring the product is useful for all members of society, in line with public interest values.

4. **Public service media recommendation systems can raise a range of well-documented ethical risks, but these will differ depending on the type of system and context of its use.** Our research found that public service media recognise a wide array of well-documented ethical risks of recommendation systems, including risks to personal autonomy, privacy, misinformation and fragmentation of the public sphere. However, the type and severity of the risks highlighted depended on which teams we spoke with, with audio-on-demand and video-on-demand teams raising somewhat different concerns to those working on news.

5. **Evaluating the risks and mitigations of recommendation systems must be done in the context of the wider product.** Addressing the risks of public service media recommendation systems should not just focus on technical fixes. Aligning product goals and other product features with public service values are just as important in ensuring recommendation systems positive contribute the experiences of audiences and to wider society.

Recommendations

Based on these key findings, we make nine recommendations for future research, experimentation and collaboration between public service media organisations, academics, funders and regulators:

1. **Define public service value for the digital age.** Recommendation systems are designed to optimise against specific objectives. However, the development and implementation of recommendation systems is happening at a time when the concept of public service value and the role of public service media organisations is under question. Unless public service media organisations are clear about their own identities and purpose, it will be difficult for them to build

effective recommendation systems. In the UK, significant work has already been done by Ofcom as well as the Department for Digital, Culture, Media and Sport's parliamentary Select Committee to identify the challenges public service media face and offer new approaches to regulation. Their recommendations must be implemented so that public service media can operate within a paradigm appropriate to the digital age and build systems that address a relevant mission.

2. **Fund a public R&D hub for recommendation systems and responsible recommendation challenges.** There is a real opportunity to create a hub for R&D of recommendation systems that are not tied to industry goals. This is especially important as recommendation systems are one of the prime use cases of behaviour modification technology but research into it is impaired by lack of access to interventional data. Therefore, as part of UKRI's National AI Research and Innovation (R&I) Programme set out in the UK AI Strategy, it should fund the development of a public research hub on recommendation technology.

3. **Publish research into audience expectations of personalisation.** There was a striking consensus in our interviews with public service media teams working on recommendations that personalisation was both wanted and expected by the audience. However, there is limited publicly available evidence underlying this belief and more research is needed. Understanding audience's views towards recommendation systems is an important part of ensuring those systems are acting in the public interest. Public service media organisations should not widely adopt recommendation systems without evidence that they are either wanted or needed by the public. Otherwise, public service media risk simply following a precedent set by commercial competitors, rather than defining a paradigm aligned to their own missions.

4. **Communicate and be transparent with audiences.** Although most public service media organisations profess a commitment to transparency about their use of recommendation systems, in practice there is little effective communication with their audiences about where and how recommendation systems are being used. Public service media should invest time and research into understanding how to usefully and honestly articulate their use of recommendation systems in ways that are meaningful to their

audiences. This communication must not be one way. There must be opportunities for audiences to give feedback and interrogate the use of the systems, and raise concerns.

5.   **Balance user control with convenience.** Transparency alone is not enough. Giving users agency over the recommendations they see is an important part of responsible recommendation. Simply giving users direct control over the recommendation system is an obvious and important first step, but it is not a universal solution. We recommend that public service media providers experiment with different kinds of options, including enabling algorithmic choice of recommendation systems and 'joint' recommendation profiles.

6.   **Expand public participation.** Beyond transparency or individual user choice and control over the parameters of the recommendation systems already deployed, users and wider society could also have greater input during the initial design of the recommendation systems and in the subsequent evaluations and iterations. This is particularly salient for public service media organisations as, unlike private companies which are primarily accountable to their customers and shareholders, public service media organisations have an obligation to serve the interests of society. Therefore, even those who are not direct consumers of content should have a say in how public service media recommendations are shaped.

7.   **Standardise metadata.** Inconsistent, poor quality metadata – an essential resource for training and developing recommendation systems – was consistently highlighted as a barrier to developing recommendation systems in public service media, particularly in developing more novel approaches that go beyond user engagement and try to create diverse feeds of recommendations. Each public service media organisation should have a central function that standardises the format, creation and maintenance of metadata across the organisation. Institutionalising the collection of metadata and making access to it more transparent across each individual organisation is an important investment in public service media's future capabilities.

8.   **Create shared recommendation system resources.** Given their limited resources and shared interests, public service media organisations should invest more heavily in creating common resources for evaluating and using recommendation systems. This

could include a shared repository for evaluating recommendation systems on metrics valued by public service media, including libraries in common coding languages.

9. **Create and empower integrated teams.** When developing and deploying recommendation systems, public service media organisations need to integrate editorial and development teams from the start. This ensures that the goals of the recommendation system are better aligned with the organisation's goals as a whole and ensure the systems augment and complement existing editorial expertise.

How to read this report

This report examines how European public service media organisations think about using automated recommendation systems for content curation and delivery. It covers the context in which recommendation systems are being deployed, why that matters, the ethical risks and evaluation difficulties posed by these systems and how public service media are attempting to mitigate these risks. It also provides ideas for new approaches to evaluation that could enable better alignment of their systems with public service values.

If you need an introduction or refresher on what recommendation systems are, we recommend starting with the 'Introducing recommendation systems' on page 31.

If you work for a public service media organisation

- We recommend the chapters on 'Stated goals and potential risks of using recommendation systems in public service media' (page 51) and 'Evaluation of recommendation systems' on page 81.

- For an understanding of how the BBC has deployed recommendation systems, see the case studies on page 104.

- For ideas on how public service media organisations can advance their responsible use of recommendation systems, see the chapter on 'Outstanding questions and areas for further research and experimentation' on page 116.

If you are a regulator of public service media

- We recommend you pay particular attention to the section on 'Stated goals and potential risks of using recommendation systems in public service media' on page 51 and 'How do public service media evaluate their recommendation systems?' on page 81.

- In addition, to understand the practices and initiatives that we believe should be encouraged within and experimented with by public service media organisations to ensure responsible and effective use of recommendation systems, see 'Outstanding questions and areas for further research and experimentation' on page 116.

If you are a regulator of online platforms

- If you need an introduction or refresher on what recommendation systems are, we recommend starting with the 'Introducing recommendation systems' on page 31. Understanding this context can help disentangle the challenges in regulating recommendation systems, by highlighting where problems arise from the goals of public service media versus the process of recommendation itself.

- To understand the issues faced by all deployers of recommendation systems, see the sections on the 'Stated goals of recommendation systems' and 'Potential risks from recommendation systems', starting on page 51.

- To better understand how these risks change due to the context and choices of public service media, relative to other online platforms, and the difficulties even organisations explicitly oriented towards public value have in auditing their own recommendation systems to determine whether they are socially beneficial, beyond simple quantitative engagement metrics, see the section on 'How these risks are viewed and addressed by public service media' starting on page 61 and the chapter on 'Evaluation of recommendation systems' starting on page 81.

If you are a funder of research into recommendation systems or a researcher interested in recommendation systems

- Public service media organisations, with mandates that emphasise social goals of universality, diversity and innovation over engagement and profit-maximising, can offer an important site of study and experimentation for new approaches to recommendation system design and evaluation. We recommend starting with the sections on 'The context of public service values and public service media' and

'Why this matters', starting on page 18, to understand the different context within which public service media organisations operate.

- Then, the sections on 'How do public service media evaluate their recommendation systems?' and 'How could evaluations be done differently?' (page 85), followed by the chapter on 'Outstanding questions and areas for further research and experimentation', starting on page 116, could provide inspiration for future research projects or pilots that you could undertake or fund.

Introduction

Scope

Recommendation systems are tools designed to sift through the vast quantities of data available online and use algorithms to guide users towards a narrower selection of material, according to a set of criteria chosen by their developers. Recommendation systems sit behind a vast array of digital experiences. 'Other users also purchased…' on Amazon or 'Watch next' on Netflix guide you to your next purchase or night on the sofa. Deliveroo will suggest what to eat, LinkedIn where to work and Facebook who your friends might be.

These practices are credited with driving the success of companies like Netflix and Spotify. But they are also blamed for many of the harms associated with the internet, such as the amplification of harmful content, the polarisation of political viewpoints (although the evidence is mixed and inconclusive)[1] and the entrenchment of inequalities.[2] Regulators and policymakers worldwide are paying increasing attention to the potential risks of recommendation systems, with proposals in China and Europe to regulate their design, features and uses.[3]

Public service media organisations are starting to follow the example of their commercial rivals and adopt recommendation systems. Like the big digital streaming service providers, they sit on huge catalogues

---

1   Cobbe, J. and Singh, J. (2019). 'Regulating Recommending: Motivations, Considerations, and Principles'. *European Journal of Law and Technology*, 10(3), pp. 8–10. Available at: https://ejlt.org/index.php/ejlt/article/view/686; Steinhardt, J. (2021). 'How Much Do Recommender Systems Drive Polarization?'. *UC Berkeley*. Available at: https://jsteinhardt.stat.berkeley.edu/blog/recsys-deepdive; Stray, J. (2021). 'Designing Recommender Systems to Depolarize', p. 2. *arXiv*. Available at: http://arxiv.org/abs/2107.04953

2   Born, G. Morris, J. Diaz, F. and Anderson, A. (2021). *Artificial intelligence, music recommendation, and the curation of culture: A white paper*, pp. 10–13. Schwartz Reisman Institute for Technology and Society. Available at: https://static1.squarespace.com/static/5ef0b24bc96ec4739e7275d3/t/60b68ccb5a371a1bcdf79317/1622576334766/Born-Morris-etal-AI_Music_Recommendation_Culture.pdf

3   See: European Union. (2022). *Digital Services Act*, Article 27. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2022:277:TOC; For details of Article 17 of the Cybersecurity Administration of China (CAC)'s Internet Information Service Algorithm Recommendation Management Regulations, see: Huld, A. (2022). 'China Passes Sweeping Recommendation Algorithm Regulations'. *China Briefing News*. Available at: https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations-effect-march-1-2022/

of news and entertainment content, and can use recommendation systems to direct audiences to particular options.

But public service media organisations face specific challenges in deploying these technologies. Recommendation systems are designed to optimise for certain objectives: a hotel's website is aiming for maximum bookings, Spotify and Netflix want you to renew your subscription. Public service media serve many functions. They have a duty to serve the public interest, not the company bottom line. They are independently financed and are controlled by, if not answerable to, the public.[4] Their mission is to inform, educate and entertain.

> Public service media are committed to values including independence, excellence and diversity.[5] They must fulfil an array of duties and responsibilities set down in legislation that often predates the digital era. How do you optimise for all that?

Developing recommendation systems for public service media is not just about finding technical fixes. It requires an interrogation of the organisations' role in democratic societies in the digital age. How do the public service values that have guided them for a century translate to a context where the internet has fragmented the public sphere and audiences are defecting to streaming services? And how can public service media use this technology in ways that serve the public interest?

These are questions that resonate beyond the specifics of public service media organisations. All public institutions that wish to use technologies for societal benefit must grapple with similar issues. And all organisations – public or private – have to deploy technologies in ways that align with their values. Asking these questions can be helpful to technologists more generally.

4   Conseil mondial de la radiotélévision. (2001). *Public broadcasting: why? how?* pp. 11–15. UNESCO Digital Library. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000124058

5   European Broadcasting Union. (2012). *Empowering Society: A Declaration on the Core Values of Public Service Media*. Available at: https://www.ebu.ch/files/live/sites/ebu/files/Publications/EBU-Empowering-Society_EN.pdf

In this report, we explore how public service media organisations are addressing the challenge of designing and implementing recommendation systems within the parameters of their organisational mission, and identify areas for further research into how they can accomplish this goal

In a context where the negative impacts of recommendation systems are increasingly apparent, public service media must tread carefully when considering their use. But there is also an opportunity for public service media to do what, historically, it has excelled at – innovating in the public interest.

> A public service approach to building recommendation systems that are both engaging and trustworthy could not only address the needs of public service media in the digital age, but provide a benchmark for scrutiny of systems more widely and create a challenge to the paradigm set by commercial operators' practices.

In this report, we explore how public service media organisations are addressing the challenge of designing and implementing recommendation systems within the parameters of their organisational mission, and identify areas for further research into how they can accomplish this goal.

While there is an extensive literature exploring public service values and a separate literature around the ethics and operational challenges of designing and implementing recommendation systems, there are still many gaps in the literature around how public service media organisations are designing and implementing these systems. Addressing that gap can help ensure that public service media organisations are better able to design these systems. With that in mind, this report explores the following questions:

- What are the values that public service media organisations adhere to? How do these differ from the goals that private-sector organisations are incentivised to pursue?

- In what contexts do public service media use recommendation systems?

- What value can recommendation systems add for public service media and how do they square with public service values?

This report explores an in-depth snapshot of the BBC's development and use of several recommendation systems, alongside an examination of the work of several other European public service media organisations

- What are the ethical risks that recommendation systems might raise in those contexts? And what challenges should different teams within public service media organisations (such as product, editorial, legal and engineering) consider?

- What are the mitigations that public service media can implement in the design, development and implementation of these systems?

In answering these questions, this report:

- provides greater clarity about the ethical challenges that developers of recommendation systems must consider when designing and maintaining these systems

- explores the social benefit of recommendation systems by examining the trade-offs between their stated goals and their potential risks

- provides examples of how public service broadcasters are grappling with these challenges, which can help inform the development of recommendation systems in other contexts.

This report focuses on European public service media organisations and in particular on the British Broadcasting Corporation (BBC) in the UK, who are project partners on this research. The BBC is the world's largest public service media organisation and has been at the forefront amongst public service broadcasters of exploring the use of recommendation systems. As the BBC has historically set precedents that other public service media have followed, it is valuable to understand its work in depth in order to draw wider lessons for the field.

In this report, we explore an in-depth snapshot of the BBC's development and use of several recommendation systems as it stood in 2021, alongside an examination of the work of several other European public service media organisations. We place these examples in the broader context of debates around 21st century public service media and use them to explore the motivations, risks and evaluation of the use of recommendation systems by public service media and their use more broadly.

The evidence for this report stems from interviews with 11 current staff from editorial, product and engineering teams, involved in recommendation systems at the BBC, along with interviews with

representatives of six other European public service broadcasters that use recommendation systems. This report also draws on a review of the existing literature on public service media recommendation systems and on interviews with experts from academia, civil society and regulation who work on the design, development, and evaluation of recommendation systems.

Although a large amount of the academic literature focuses on the use of recommendations in news provision, we look at the full range of public service media content, as we found more of the advanced implementations of recommendation systems lie in other domains. We have drawn on published research about recommendation systems from commercial platforms, however, internal corporate studies are unavailable to independent researchers and our requests to interview both researchers and corporate representatives of platforms were unsuccessful.

Background

In this chapter, we set out the context for the rest of the report. We outline the history and context of public service media organisations, what recommendation systems are and how they are approached by public service media organisations, and what external and internal processes and constraints govern their use.

The context of public service values and public service media

The use of recommendation systems in public service media is informed by their history, values and remit, their governance and the landscape in which they operate. In this section we situate the deployment of recommendation systems in this context.

Broadly, public service media are independent organisations that have a mission to inform, educate and entertain. Their values are rooted in the founding vision for public service media organisations a century ago and remain relevant today, codified into regulatory and governance frameworks at organisational, national and European levels. However the values that public service media operate under are inherently qualitative and, even with the existence of extensive guidelines, are interpreted through the daily judgements of public service media staff and the mental models and institutional culture built up over time.

Although public service media have been resilient to change, they currently face a trio of challenges:

1. Losing audiences to online digital content providers including Netflix, Amazon, YouTube and Spotify.

2. Budget cuts and outdated regulation, framed around analogue broadcast commitments, hampering their ability to respond to technological change.

3. Populist political movements undermining their independence.

Public service media are independent media organisations financed by and answerable to the publics they serve.[6] Their roots lie in the 1920s technological revolution of radio broadcasting when the BBC was established as the world's first public service broadcaster, funded by a licence fee, and with the ambition to 'bring the best of everything to the greatest number of homes'.[7] Other national broadcasters were soon founded across Europe and also adopted the BBC's mission to 'inform, educate and entertain'. Although there are now public service media organisations in almost every country in the world, this report focuses on European public service media, which share comparable social, political and regulatory developments and therefore a similar context when considering the implementation of recommendation systems.

Public service media organisations have come to play an important institutional role within democratic societies in Europe, creating a bulwark against the potential control of public opinion either by the state or by particular interest groups.[8] The establishment of public service broadcasters for the first time created a universally accessible public sphere where, in the words of the BBC's founding chairman Lord Reith, 'the genius and the fool, the wealthy and the poor listen simultaneously'. They aimed to forge a collective experience, 'making the nation as one man'.[9] At the same time public service media are expected to reflect the diversity of a nation, enabling the wide representation of perspectives in a democracy, as well as giving people sufficient information and understanding to make decisions on issues of public importance. These two functions create an inherent tension between public service media as an agonistic space where different viewpoints compete and a consensual forum where the nation comes together.

Public service values

The founding vision for public service media has remained within the DNA of organisations as their public service values – often called

6   Conseil mondial de la radiotélévision. (2001). *Public broadcasting: why? how?* pp. 11–15. UNESCO Digital Library.
       Available at: https://unesdoc.unesco.org/ark:/48223/pf0000124058
7   BBC. (2022). *The BBC Story - 1920s factsheet*. Available at: http://downloads.bbc.co.uk/historyofthebbc/1920s.pdf
8   Tambini, D. (2021). 'Public service media should be thinking long term when it comes to AI'. *Media@LSE*. Available at:
       https://blogs.lse.ac.uk/medialse/2021/05/12/public-service-media-should-be-thinking-long-term-when-it-comes-to-ai/
9   Higgins, C. (2014). 'What can the origins of the BBC tell us about its future?'. *The Guardian*. Available at:
       https://www.theguardian.com/media/2014/apr/15/bbc-origins-future

Reithian principles, in reference to the influence of the BBC's founding chairman.

The European Broadcasting Union (EBU), the membership organisation for public service media in Europe, has codified the public service mission into six core values: **universality**, **independence**, **excellence**, **diversity**, **accountability** and **innovation**, and member organisations commit to strive to uphold these in practice.[10]

Public service value	Meaning
Universality	• reach all segments of society, with no-one excluded • share and express a plurality of views and ideas • create a public sphere, in which all citizens can form their own opinions and ideas, aiming for inclusion and social cohesion • multi-platform • accessible for everyone • enable audiences to engage and participate in a democratic society.
Independence	• trustworthy content • act in the interest of audiences • completely impartial and independent from political, commercial and other influences and ideologies • autonomous in all aspects of the remit such as programming, editorial decision-making, staffing • independence underpinned by safeguards in law.
Excellence	• high standards of integrity professionalism and quality; create benchmarks within the media industries • foster talent • empower, enable and enrich audiences • audiences are also participants.
Diversity	• reflect diversity of audiences by being diverse and pluralistic in the genres of programming, the views expressed, and the people employed • support and seek to give voice to a plurality of competing views – from those with different backgrounds, histories and stories. Help build a more inclusive, less fragmented society.

10   European Broadcasting Union. (2012). *Empowering Society: A Declaration on the Core Values of Public Service Media*. Available at: https://www.ebu.ch/files/live/sites/ebu/files/Publications/EBU-Empowering-Society_EN.pdf

Accountability	• listen to audiences and engage in a permanent and meaningful debate
	• publish editorial guidelines. Explain. Correct mistakes. Report on policies, budgets, editorial choices
	• be transparent and subject to constant public scrutiny
	• be efficient and managed according to the principles of good governance.
Innovation	• enrich the media environment
	• be a driving force of innovation and creativity
	• develop new formats, new technologies, new ways of connectivity with audiences
	• attract, retain and train our staff so that they can participate in and shape the digital future, serving the public.

As well as signing up to these common values, each individual public service media organisation has its own articulation of its mission, purpose and values, often set out as part of its governance.[11] Ultimately these will align with those described by the EBU but may use different terms or have a different emphasis. Policymakers and practitioners operating at a national level are more likely to refer to these specific expressions of public values. The overarching EBU values are often referenced in academic literature as the theoretical benchmark for public service values.

In the case of the BBC, the Royal Charter between the Government and the BBC is agreed for a 10 year period.[12]

---

11   Statutory governance of public service media also varies from country to country and reflects national political and regulatory norms. The BBC is regulated by the independent broadcasting regulator Ofcom. The European Union's revised Audio Visual Service Directive requires member states to have an independent regulator but this can take different forms. See: European Commission. (2018). *Digital Single Market: updated audiovisual rules*.
Available at: https://ec.europa.eu/commission/presscorner/detail/en/MEMO_18_4093. For example, France has a central regulator, the Conseil Supérieur de l'Audiovisuel. But in Germany, although public service media objectives are defined in the constitution, oversight is provided by a regional broadcasting council, Rundfunkrat, reflecting the country's federal structure. In Belgium too, regulation is devolved to two separate councils representing the country's French and Flemish speaking regions.

12   BBC. (2017). 'Mission, values and public purposes'. Available at:
https://www.bbc.com/aboutthebbc/governance/bbc.com/aboutthebbc/governance/mission/. For comparison, ARD, the German public service media organisation articulates its values as: 'Participation, Independence, Quality, Diversity, Localism, Innovation, Value Creation, Responsibility'. See: ARD. (2021). *Die ARD - Unser Beitrag zum Gemeinwohl*. Available at:
https://www.ard.de/die-ard/was-wir-leisten/ARD-Unser-Beitrag-zum-Gemeinwohl-Public-Value-100

The BBC: governance and values

Mission: to act in the public interest, serving all audiences through the provision of impartial, high-quality and distinctive output and services which inform, educate and entertain.

Public purposes:

1. To provide impartial news and information to help people understand and engage with the world around them.

2. To support learning for people of all ages.

3. To show the most creative, highest quality and distinctive output and services.

4. To reflect, represent and serve the diverse communities of all of the United Kingdom's nations and regions and, in doing so, support the creative economy across the United Kingdom.

5. To reflect the United Kingdom, its culture and values to the world.

Additionally, the BBC has its own set of organisational values that are not part of the governance agreement but that 'represent the expectations we have for ourselves and each other, they guide our day-to-day decisions and the way we behave':

- *Trust: Trust is the foundation of the BBC – we're independent, impartial and truthful.*
- *Respect: We respect each other – we're kind, and we champion inclusivity.*
- *Creativity: Creativity is the lifeblood of our organisation.*
- *Audiences: Audiences are at the heart of everything we do.*
- *One BBC: We are One BBC – we collaborate, learn and grow together.*
- *Accountability: We are accountable and deliver work of the highest quality.*

These kinds of regulatory requirements and values are then operationalised internally through organisations' editorial guidelines which again will vary from organisation to organisation, depending on the norms and expectations of their publics. Guidelines can be extensive and their aim is to help teams put public service values into practice. For example, the current BBC guidelines run to 220 pages, covering everything from how to run a competition, to reporting on wars and acts of terror.

Nonetheless, such guidelines leave a lot of room for interpretation. Public service values are, by their nature, qualitative and difficult to measure objectively. For instance, consider the BBC guidelines on impartiality – an obligation that all regulated broadcasters in the UK must uphold – and over which the BBC has faced intense scrutiny:

> 'The BBC is committed to achieving due impartiality in all its output. This commitment is fundamental to our reputation, our values and the trust of audiences. The term "due" means that the impartiality must be adequate and appropriate to the output, taking account of the subject and nature of the content, the likely audience expectation and any signposting that may influence that expectation.'

> 'Due impartiality usually involves more than a simple matter of 'balance' between opposing viewpoints. We must be inclusive, considering the broad perspective and ensuring that the existence of a range of views is appropriately reflected. It does not require absolute neutrality on every issue or detachment from fundamental democratic principles, such as the right to vote, freedom of expression and the rule of law. We are committed to reflecting a wide range of subject matter and perspectives across our output as a whole and over an appropriate timeframe so that no significant strand of thought is under-represented or omitted.'

It's clear that impartiality is a question of judgement and may not even be expressed in a single piece of content but over the range of BBC output over a period of time. In practice, teams internalise these expectations and make decisions based on institutional culture and internal mental models of public service value, rather than continually checking the editorial guidelines or referencing any specific public values matrix.[13]

---

13   Mazzucato, M., Conway, R., Mazzoli, E., Knoll E. and Albala, S. (2020). *Creating and measuring dynamic public value at the BBC*, p.22. UCL Institute for Innovation and Public Purpose. Available at: https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/final-bbc-report-6_jan.pdf

How public service media differ from other media organisations

Public service media are answerable to the publics they serve.[14] They should be independent from both government influence and from the influence of commercial owners. They operate to serve the public interest.

Commercial media, however, serve the interests of their owners or shareholders. Success for Netflix for example is measured in numbers of subscribers which then translates into revenues.[15]

The activities of commercial media are nonetheless limited by regulation. In the UK the independent regulator Ofcom's Broadcasting Code requires all broadcasters (not just public service media) to abide by principles such as fairness and impartiality.[16] Russia Today for example has been investigated for allegedly misleading reporting on the conflict in Ukraine.[17] Streaming services are subject to more limited regulation which covers child protection, incitement to hatred and product placement,[18] while the press – both online and in print – are largely lightly self-regulated through the Independent Press Standards Organisation, with some publications regulated by IMPRESS.[19]

However, public service media have extensive additional obligations, amongst others to 'meet the needs and satisfy the interests of as many different audiences as practicable' and 'reflect the lives and concerns of different communities and cultural interests and traditions within the United Kingdom, and locally in different parts of the United Kingdom'.[20]

These regulatory systems vary from country to country but hold broadly the same characteristics. In all cases, the public service remit entails

---

14   Not all public service media are publicly funded. Channel 4 in the UK for example is financed through advertising but owned by the public (although the UK Government has opened a consultation on privatisation).

15   Circulation and profits for print media have declined in recent years but in some cases promote their proprietors' interests through political influence – for instance the Murdoch-owned *Sun* in the UK or the Axel Springer-owned *Bild Zeitung* in Germany.

16   Ofcom. (2020). *The Ofcom Broadcasting Code (with the Cross-promotion Code and the On Demand Programme Service Rules)*. Available at: https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code

17   Ofcom. (2022). 'Ofcom launches 15 investigations into RT'. Available at: https://www.ofcom.org.uk/news-centre/2022/ofcom-launches-investigations-into-rt

18   Ofcom. (2021). *Guide to video on demand*. Available at: https://www.ofcom.org.uk/tv-radio-and-on-demand/advice-for-consumers/television/video-on-demand

19   Independent Press Standards Organisation (IPSO). (2022). 'What we do'. Available at: https://www.ipso.co.uk/what-we-do/; IMPRESS. 'Regulated Publications'. Available at: https://impress.press/regulated-publications/

20   UK Government. *Communications Act 2003*, section 265. Available at: https://www.legislation.gov.uk/ukpga/2003/21/section/265

far greater duties than in the private sector and broadcasters are more heavily regulated than digital providers.

These obligations are also framed in terms of public or societal benefit. This means public service media are striving to achieve societal goals that may not be aligned with a pure maximisation of profits, while commercial media pursue interests more aligned with revenue and the interests of their shareholders.

Nonetheless, public service media face scrutiny about how well they meet their objectives and have had to create proxies for these intangible goals to demonstrate their value to society.

> *'[Public service media] is fraught today with political contention. It must justify its existence and many of its efforts to governments that are sometimes quite hostile, and to special interest groups and even competitors. Measuring public value in economic terms is therefore a focus of existential importance; like it or not diverse accountability processes and assessment are a necessity.'* [21]

In practice this means public service media organisations measure their services against a range of hard metrics, such as audience reach and value for money, as well as softer measures like audience satisfaction surveys.[22] In the mid-2000s the BBC developed a public value test to inform strategic decisions that has since been adopted as a public interest test which remains part of the BBC's governance. Similar processes have been created in other public service media systems, such as the 'Three Step Test' in German broadcasting.[23] These methods have their own limitations, drawing public media into a paradigm of cost-benefit analysis and market fixing, rather than articulating wider values to individuals, society and industry.[24]

---

21   Lowe, G. and Martin, F. (eds.). (2014). *The Value and Values of Public Service Media*.

22   BBC. (2021). *BBC Annual Plan 2021-22*, Annex 1. Available at:
      http://downloads.bbc.co.uk/aboutthebbc/reports/annualplan/annual-plan-2021-22.pdf

23   The 12th Inter-State Broadcasting Treaty, the regulatory framework for public service and commercial broadcasting across Germany's federal states, introduced a three-step test for assessing whether online services offered by public service broadcasters met their public service remit. Under the three-step test, the broadcaster needs to assess: first, whether a new or significantly amended digital service satisfies the democratic, social and cultural needs of society; second, whether it contributes to media competition from a qualitative point of view and; third, the associated financial cost. See: Institute for Media and Communication Policy. (2009). *Drei-Stufen-Test*. Available at: http://medienpolitik.eu/drei-stufen-test/

24   Mazzucato, M., Conway, R., Mazzoli, E., Knoll E. and Albala, S. (2020). *Creating and measuring dynamic public value at the BBC*, p.22. UCL Institute for Innovation and Public Purpose. Available at:
      https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/final-bbc-report-6_jan.pdf

This does not mean commercial media are devoid of values. Spotify for example says its mission 'is to unlock the potential of human creativity – by giving a million creative artists the opportunity to live off their art and billions of fans the opportunity to enjoy and be inspired by it',[25] while Netflix's organisational values are judgment, communication, curiosity, courage, passion, selflessness, innovation, inclusion, integrity and impact.[26] Commercial media are also sensitive to issues that present reputational risk, for instance the outcry over Joe Rogan's Spotify podcast propagating disinformation about COVID-19 or Jimmy Carr's joke about the Holocaust.[27]

However, commercial media harness values in service of their business model, whereas for public service media the values themselves are the organisational objective. Therefore, while the ultimate goal of a commercial media organisation is quantitative (revenue) the ultimate goal of public service media is qualitative (public value) – even if this is converted into quantitative proxies.

This difference between public and private media companies is fundamental in how they adopt recommendation systems. We discuss this further later in the report when examining the objectives of using recommendation systems.

Current challenges for public service media

Since their inception, public service media and their values have been tested and reinterpreted in response to new technologies.

The introduction of the BBC Light Programme in 1945, a light entertainment alternative to the serious fare offered by the BBC Home Service, challenged the principle of universality (not everyone was listening to the same content at the same time) as well as the balance between the mission to inform, educate and entertain (should public service broadcasting give people what they want or

25   Spotify.(2022). 'About Spotify'. Available at: https://newsroom.spotify.com/company-info/

26   Netflix. (2022). 'Netflix Culture'. Available at: https://jobs.netflix.com/culture

27   Silberling, A. (2022). 'Spotify adds COVID-19 content advisory'. *TechCrunch*. Available at: https://social.techcrunch.com/2022/03/28/spotify-covid-19-content-advisory-joe-rogan/; Jackson, S. (2022). 'Jimmy Carr condemned by Nadine Dorries for "shocking" Holocaust joke about travellers in Netflix special His Dark Material'. *Sky News*. Available at: https://news.sky.com/story/jimmy-carr-condemned-for-disturbing-holocaust-joke-about-travellers-in-netflix-special-his-dark-material-12533148

what they need?). The arrival of the video recorder, and then new channels and platforms, gave audiences an option to opt out of the curated broadcast schedule –where editors determined what should be consumed. While this enabled more and more personalised and asynchronous listening and viewing, it potentially reduced exposure to the serendipitous and diverse content that is often considered vital to the public service remit.[28]

The arrival and now dominance of digital technologies comes amid a collision of simultaneous challenges which, in combination, may be existential.

Audience

Public service media have always had a hybrid role. They are obliged to serve the public simultaneously as citizens and consumers.[29] Their public service mandate requires them to produce content and serve audiences that the commercial market does not provide for. At the same time, their duty to provide a universal service means they must aim to reach a sizeable mainstream audience and be active participants in the competitive commercial market.

Although people continue to use and value public service media, the arrival of streaming services such as Netflix, Amazon and Spotify, as well as the availability of content on YouTube, has had a massive impact on public service media audience share.

In the UK, the COVID-19 pandemic has seen people return to public service media as a source of trusted information, and with more time at home they have also consumed more public service content.[30]

But lockdowns also supercharged the uptake of streaming. By September 2020, 60% of all UK households subscribed to an on-demand service, up from 49% a year earlier. Just under half (47%) of all adults who go online now consider online services to be their

28   van Es, K. F. (2017). 'An Impending Crisis of Imagination : Data-Driven Personalization in Public Service Broadcasters'. *Media@LSE*. Available at: https://dspace.library.uu.nl/handle/1874/358206

29   BBC Trust. (2012). *BBC Trust assessment processes Guidance document*. Available at: http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/about/how_we_govern/pvt/assessment_processes_guidance.pdf

30   BBC. (2021). *Annual Plan 2021-22*. Available at: http://downloads.bbc.co.uk/aboutthebbc/reports/annualplan/annual-plan-2021-22.pdf

main way of watching TV and films, rising to around two-thirds (64%) among 18–24 year olds.[31]

Public service media are particularly concerned about their failure to reach younger audiences.[32] Although this group still encounters public service media content, they tend to do so on external services: younger viewers (16–34 year olds) are more likely to watch BBC content on subscription video-on-demand (SVoD) services rather than through BBC iPlayer (4.7 minutes per day on SVoD vs. 2.5 minutes per day on iPlayer).[33] They are not necessarily aware of the source of the content and do not create an emotional connection with the public service media as a trusted brand. Meanwhile, platforms gain valuable audience insight data through this consumption which they do not pass onto the public service media organisations.[34]

Regulation

Legislation has not kept pace with the rate of technological change. Public service media are trying to grapple with the dynamics of the competitive digital landscape on stagnant or declining budgets, while continuing to meet their obligations to provide linear TV and radio broadcasting to a still substantial legacy audience.

The UK broadcasting regulator Ofcom published recommendations in 2021, repeating its previous demands for an urgent update to the public service media system to make it sustainable for the future. These include modernising the public service objectives, changing licences to apply across broadcast and online services and allowing greater flexibility in commissioning across platforms.[35]

31   Ofcom. (2021). *Small Screen: Big Debate – Recommendations to Government on the future of Public Service Media*. Available at: https://www.smallscreenbigdebate.co.uk/__data/assets/pdf_file/0023/221954/statement-future-of-public-service-media.pdf

32   Lowe, G.F. and Maijanen, P. (2019). 'Making sense of the public service mission in media: youth audiences, competition, and strategic management'. *Journal of Media Business Studies*. doi: 10.1080/16522354.2018.1553279; Schulz, A., Levy, D. and Nielsen, R.K. (2019). 'Old, Educated, and Politically Diverse: The Audience of Public Service News', pp. 15–19, 29–30. *Reuters Institute for the Study of Journalism*. Available at: https://reutersinstitute.politics.ox.ac.uk/our-research/old-educated-and-politically-diverse-audience-public-service-news

33   Ofcom. (2021). *Small Screen: Big Debate – Recommendations to Government on the future of Public Service Media*. Available at: https://www.smallscreenbigdebate.co.uk/__data/assets/pdf_file/0023/221954/statement-future-of-public-service-media.pdf

34   House of Commons Digital, Culture, Media and Sport Committee. (2021). *The future of public service broadcasting*, HC 156. Available at: https://publications.parliament.uk/pa/cm5801/cmselect/cmcumeds/156/156.pdf

35   Ofcom. (2021). S*mall Screen: Big Debate – Recommendations to Government on the future of Public Service Media*. Available at: https://www.smallscreenbigdebate.co.uk/__data/assets/pdf_file/0023/221954/statement-future-of-public-service-media.pdf

The Digital, Culture, Media and Sport Select Committee of the House of Commons has also demanded regulatory change. It warned that 'hurdles such as the Public Interest Test inhibit the ability of [public service broadcasters] to be agile and innovate at speed in order to compete with other online services' and that the core principle of universality would be threatened unless public service media were better able to attract younger audiences.[36]

Although there has been a great deal of activity around other elements of technology regulation, particularly the Online Safety Bill in the UK and the Digital Services Act in the European Union, the regulation of public service media has not been treated with the same urgency. There is so far no Government white paper for a promised Media Bill that would address this in the UK and the European Commission's proposals for a European Media Freedom Act are in the early stages of consultation.[37]

Political context

Public service media have always been a political battleground and have often had fractious relationships with the government of the day. But the rise of populist political movements and governments has created new fault lines and made public service media a battlefield in the culture wars. The Polish and Hungarian Governments have moved to undermine the independence of public service media, while the far-right AfD party in eastern Germany refused to approve funding for public broadcasting.[38] In the UK, the Government has frozen the licence fee for two years and has said future funding arrangements are 'up for discussion'. It has also been accused of trying to appoint an ideological ally to lead the independent media regulator Ofcom. Elsewhere in Europe, journalists from public service media have been attacked by anti-immigrant and COVID-denial protesters.[39]

At the same time, public service media are criticised as unrepresentative of the publics they are supposed to serve. In the UK, both the BBC and

36   House of Commons Digital, Culture, Media and Sport Committee. (2021). *The future of public service broadcasting*, HC 156. Available at: https://publications.parliament.uk/pa/cm5801/cmselect/cmcumeds/156/156.pdf

37   European Commission. (2022). 'European Media Freedom Act: Commission launches public consultation'. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_85

38   The Economist. (2021). 'Populists are threatening Europe's independent public broadcasters'. Available at: https://www.economist.com/europe/2021/04/08/populists-are-threatening-europes-independent-public-broadcasters

39   The Economist. (2021).

Channel 4 have attempted to address this by moving parts of their workforce out of London.[40] As social media has removed traditional gatekeepers to the public sphere, there is less acceptance of and deference towards the judgement of media decision-makers. In a fragmented public sphere, it becomes harder for public service media to 'hold the ring' – on issues like Brexit, COVID-19, race and transgender rights, public service media find themselves distrusted by both sides of the argument.

Although the provision of information and educational resources through the COVID-19 pandemic has given public service media a boost, both in audiences and in levels of trust, they can no longer take their societal value or even their continued existence for granted.[41] Since the arrival of the internet, their monopoly on disseminating real-time information to a wide public has been broken and so their role in both the media and democratic landscape is up for grabs.[42] For some, this means public service media is redundant.[43] For others, its function should now be to uphold national culture and distinctiveness in the face of the global hegemony of US-owned platforms.[44]

The Institute for Innovation and Public Purpose has proposed reimagining the BBC as a 'market shaper' rather than a market fixer, based on a concept of dynamic public value,[45] while the Media Reform Coalition calls for the creation of a Media Commons of independent, democratic and accountable media organisations, including a People's BBC and Channel 4.[46] The wide range of ideas in play demonstrates how open the possible futures of public service media could be.

40   The Sutton Trust. (2019). *Elitist Britain*, pp. 40–42. Available at: https://www.suttontrust.com/our-research/elitist-britain-2019/; Friedman, S. and Laurison, D. (2019). 'The class pay gap: why it pays to be privileged'. *The Guardian*. Available at: https://www.theguardian.com/society/2019/feb/07/the-class-pay-gap-why-it-pays-to-be-privileged

41   BBC. (2021). *Annual Plan 2021-22*. Available at: http://downloads.bbc.co.uk/aboutthebbc/reports/annualplan/annual-plan-2021-22.pdf

42   Interview with Jannick Kirk Sørensen, Associate Professor in Digital Media, Aalborg University (2021).

43   Booth, P. (2020). *New Vision: Transforming the BBC into a subscriber-owned mutual*. Institute of Economic Affairs. Available at: https://iea.org.uk/publications/new-vision

44   Department for Digital, Culture, Media & Sport and John Whittingdale OBE MP. (2021). *John Whittingdale's speech to the RTS Cambridge Convention 2021*. UK Government. Available at: https://www.gov.uk/government/speeches/john-whittingdales-speech-to-the-rts-cambridge-convention-2021

45   Mazzucato, M., Conway, R., Mazzoli, E., Knoll E. and Albala, S. (2020). *Creating and measuring dynamic public value at the BBC*, p.22. UCL Institute for Innovation and Public Purpose. Available at: https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/final-bbc-report-6_jan.pdf

46   Grayson, D. (2021). *Manifesto for a People's Media*. Media Reform Coalition. Available at: https://drive.google.com/file/u/1/d/1_6GeXiDR3DGh1sYjFI_hbgV9HfLWzhPi/view?usp=embed_facebook

In this section, we briefly outline how recommendation systems operate and how they are used in practice by European public service media.

Introducing recommendation systems

The main steps in the development of a recommendation: user engagement with the platform, data gathering, algorithmic analysis and recommendation generation.
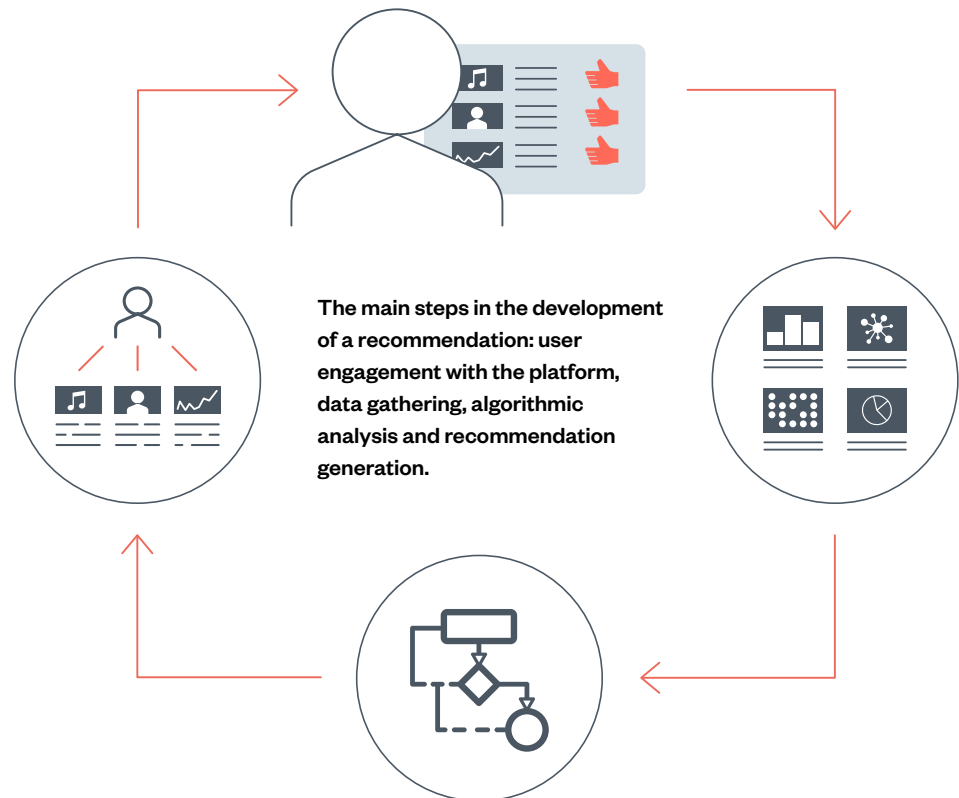
Day-to-day, we might turn to friends or family for their recommendations when it comes to decisions large and small. From dining out and entertainment, to big purchases. We might also look at expert reviews. But in the last decade, there has been a critical change in where recommendations come from and how they're used. Recommendations have now become a pervasive feature of the digital products we use.

Recommendation systems are a type of software that filter information based on contextual data and according to criteria set by its designers. In this section, we briefly outline how recommendation systems operate and how they are used in practice by European public service media. At least a quarter of European public service media have begun deploying recommendation systems. They are mainly used on video platforms but they are only applied on small sections of services – the vast majority of public service content continues to be manually curated by editors.

In media organisations, producers, editors and journalists have always made implicit and explicit decisions about what to give prominence to, from what stories to tell and what programmes to commission, to – just as importantly – how those stories are presented. Deciding what makes the front page, what gets prime time, what makes top billing on the evening news – these are all acts of recommendation. For some, the entire institution is a system for recommending content to their audiences.

Public service media organisations are starting to automate these decisions by using recommendation systems.

Recommendation systems are context-driven information filtering systems. They don't use explicit search queries from the user (unlike search engines) and instead rank content based only on contextual information.[47] This can include:

- the item being viewed, e.g. the current webpage, the article being read, the video that just finished playing etc.
- the item being filtered and recommended, e.g. the length of the content, when the content was published, characteristics of the content, e.g. drama, sport, news - often described as metadata about the content
- the users, e.g. their location or language preferences, their past interactions with the recommendation system etc.
- the wider environment, e.g. the time of day.

Examples of well-known products utilising recommendation systems include:

- Netflix's homepage
- Spotify's auto-generated playlists and auto-play features
- Facebook's 'People You May Know' and 'News Feed'
- YouTube's video recommendations

47   Tennenholtz, M. and Kurland, O. (2019). 'Rethinking Search Engines and Recommendation Systems: A Game Theoretic Perspective'. *Communications of the ACM, December 2019,* 62(12), pp. 66–75. Available at: https://cacm.acm.org/magazines/2019/12/241056-rethinking-search-engines-and-recommendation-systems/fulltext; Jannach, D. and Adomavicius, G. (2016), 'Recommendations with a Purpose'. *RecSys '16: Proceedings of the 10th ACM Conference on Recommender Systems*, pp7–10. Available at: https://doi.org/10.1145/2959100.2959186; Jannach, D., Zanker, M., Felfernig and Friedrich, G. (2010). *Recommender Systems: An Introduction.* Cambridge University Press. doi: 10.1017/CBO9780511763113; Ricci, F., Rokach, L. and Shapira, B. (2015). *Recommender Systems Handbook.* Springer New York: New York. doi: 10.1007/978-1-4899-7637-6

In this report, we have chosen to focus on recommendation systems rather than search engines as the context-driven rather than query-driven approach of recommendation systems is much more analogous to traditional human editorial judgment and content curation.

- TikTok's 'For You' page
- Amazon's 'Recommended For You', 'Frequently Bought Together', 'Items Recently Viewed', 'Customers Who Bought This Item Also Bought', 'Best-Selling' etc.[48]
- Tinder's swiping page[49]
- LinkedIn's 'Recommend for you' jobs page
- Deliveroo or UberEats' 'recommended' sort for restaurants.

Recommendation systems and search engines

It is worth acknowledging the difference between recommendation systems and **search engines**, which can be thought of as **query-driven information filtering systems**. They filter, rank and display webpages, images and other items primarily in response to a query from a user (such as Google searching for 'restaurants near me'). This is then often combined with the contextual information mentioned above. Google Search is the archetypal search engine in most Western countries but other widely used search engines include Yandex, Baidu and Yahoo. Many public service media organisations offer a query-driven search feature on their services that enables users to search for news stories or entertainment content.

In this report, we have chosen to focus on recommendation systems rather than search engines as the context-driven rather than query-driven approach of recommendation systems is much more analogous to traditional human editorial judgment and content curation.

Broadly speaking, recommendation systems take a series of inputs, filter and select which ones are most important, and produce an output (the recommendation). The inputs and outputs of recommendation systems are subject to content moderation (in which the pool of content is pre-screened and filtered) and curation (in which content is selected, organised and presented).

48   Singh, S. (2020). *Why Am I Seeing This? – Case study: Amazon*. New America. Available at:
     https://www.newamerica.org/oti/reports/why-am-i-seeing-this

49   Liu, S. (2017). 'Personalized Recommendations at Tinder' [presentation]. Available at:
     https://www.slideshare.net/SessionsEvents/dr-steve-liu-chief-scientist-tinder-at-mlconf-sf-2017

This starts by deciding what to input into the recommendation system. The pool of content to draw from is often dictated by the nature of the platform itself, such as activity from your friends, groups, events, etc. alongside adverts, as in the case of Facebook. In the case of public service media, the pool of content is often their back catalogue of audio, video or news content.

This content will have been moderated in some way before it reaches the recommendation system, either manually by human moderators or editors, or automatically through software tools. On Facebook, this means attempts to remove inappropriate user content, such as misinformation or hate speech, from the platform entirely, according to moderation guidelines. For a public service media organisation, this will happen in the commissioning and editing of articles, radio programmes and TV shows by producers and editorial teams.

The pool of content will then be further curated as it moves through the recommendation system, as certain pieces of content might be deemed appropriate to publish but not to recommend in a particular context, e.g. Facebook might want to avoiding recommending you posts in languages you don't speak. In the case of public service media, this generally takes the form of business rules, which are editorial guidelines implemented directly into the recommendation system.

Some business rules apply equally across all users and further constrain the set of content that the system recommends content from, such as only selecting content from the past few weeks. Other rules apply after individual user recommendations have been generated and filter those recommendations based on specific information about the user's context, such as not recommending content the user has already consumed.

For example, below are business rules that were implemented in BBC Sounds' Xantus recommendation system, as of summer 2021:[50]

Non-personalised business rules	Personalised business rules
Recency	Already seen items
Availability	Local radio (if not consumed previously)
Excluded 'master brands', e.g., particular radio channels[51]	Specific language (if not consumed previously)
Excluded genres	Episode picking from a series
	Diversification (1 episode per brand/series)

How different types of recommendation systems work

Not all recommendation systems are the same. One major difference relates to what categories of items a system is filtering and curating for. This can include, but isn't limited to:

- content, e.g. news articles, comments, user posts, podcasts, songs, short-form video, long-form video, movies, images etc. or any combination of these content types
- people, e.g. dating app profiles, Facebook profiles, Twitter accounts etc.
- metadata, e.g. the time, data, location, category etc. of a piece of content or the age, gender, location etc. of a person.

---

50   Note that the business rules are subject to change, and so the rules given here are intended to be an indicative example only, representing a snapshot of practice at one point in time. See: Al-Chueyr Martins, T. (2021). 'From an idea to production: the journey of a recommendation engine' [presentation recording]. *MLOps London.* Available at: https://www.youtube.com/watch?v=dFXKJZNVgw4

51   Smethurst, M. (2014). *Designing a URL structure for BBC programmes.* Available at: https://smethur.st/posts/176135860

In this report, we mainly focus on:

1. **Media content recommendation systems:** these systems rank and display pieces of media content, e.g. news articles, podcasts, short-form videos, radio shows, television shows, movies etc. to users of news websites, video-on-demand and streaming services, music and podcast apps etc.

2. **Media content metadata recommendation systems:** these rank and display suggestions for information to classify pieces of media content, e.g. genre, people or places which appear in the piece of media, or other tags, to journalists, editors or other members of staff at media organisations.

Another important distinction between applications of recommendation systems is the role of the provider in choosing which set of items the recommendation system is applied to. There are three categories of use for recommendation systems:

1. **Open recommending:** The recommendation system operates primarily on items that are generated by users of the platform, or otherwise indiscriminately automatically aggregated from other sources, without the platform curating or individually approving the items. Examples include YouTube, TikTok's 'For You' page, Facebook's News Feed' and many dating apps.

2. **Curated recommending:** The recommendation system operates on items which are curated, approved or otherwise editorialised by the platform operating the recommendation system. These systems still primarily rely on items generated by external sources, sometimes blended with items produced by the platform. Often these external items will come in the form of licensed or syndicated content such as music, films, TV shows, etc. rather than user-generated items. Examples include Netflix, Spotify and Disney+.

3. **Closed recommending:** The recommendation system operates exclusively on items generated or commissioned by the platform operating the recommendation system. Examples include most recommendation systems used on the website of news organisations.

Lastly, there are different types of technical approaches that a recommendation system may use to sort and filter content. The approaches detailed below are not mutually exclusive and can be combined in recommendation systems in particular contexts:

Type of filtering	Example	What does it do?
Collaborative filtering	'Customers Who Bought This Item Also Bought' on Amazon	The system recommends items to users based on the past interactions and preferences of other users who are classified as having similar past interactions and preferences. These patterns of behaviour from other users are used to predict how the user seeing the recommendation would rate new items. Those item rating predictions are used to generate recommendations of items that have a high level of similarity with content previously popular with similar users.
Matrix factorisation	Netflix's 'Watch Next' feature	A subclass of collaborative filtering, this method codifies users and items into a small set of categories based on all the user ratings in a system. When Netflix recommends movies, a user may be codified by how much they like action, comedy, etc. and a movie might be codified by how much it fits into these genres. This codified representation can then be used to guess how much a user will like a movie they haven't seen before, based on whether these codified summaries 'match'.
Content-based filtering	Netflix's 'Action Movies' list	These methods recommend items based on the codified properties of the item stored in the database. If the profile of items a user likes mostly consists of action films, the system will recommend other items that are tagged as action films. The system does not draw on user data or behaviour to make recommendations.

Of these typologies, the public service media that we surveyed only use closed recommendation systems as they are applying recommendations to content they have commissioned or produced. However, we found examples of public service media using all types of filtering approaches: collaborative filtering, content-based filtering and hybrid recommendation systems.

How do European public service media organisations use recommendation systems?

The use of recommendation systems is common but not ubiquitous among public service media organisations in Europe. As of 2021, at least a quarter of European Broadcasting Union (EBU) member organisations were using recommendation systems on at least one of their content delivery platforms.[52] Video-on-demand platforms are the most common use case for recommendation systems, followed by audio-on-demand and news content. As well as these public-facing recommendation systems, some public service media also use recommendation systems for internal-only purposes, such as systems that assist journalists and producers with archival research.[53]

Figure 1: Recommendation system use by European public service media by platform (EBU, 2020)

Platform on which public service media offers personalised recommendations	Number of European Broadcasting Union member organisations	Examples
Video-on-demand	At least 18	BBC iPlayer
Audio-on-demand	At least 10	BBC Sounds, ARD Audiothek
News content	At least 7	VRT NWS app

Among the EBU member organisations which reported using recommendation systems in a 2020 survey, recommendations were displayed:

- in a dedicated section on the on-demand homepage (by at least 16 organisations)
- in the player as 'play next' suggestions (by at least 10 organisations)
- as 'top picks' on the on-demand homepage (by at least 9 organisations).

---

52   See Annex 1 for more details

53   Interview with Ben Fields, Lead Data Scientist, Digital Publishing, BBC (2021).

Even among organisations that have adopted recommendation systems, their use remains very limited. NPO in the Netherlands was the only organisation we encountered that aims to have a fully algorithmically driven homepage on its main platform. In most cases, the vast majority of content remains under human editorial control, with only small sub-sections of the interface offering recommended content.

> As editorial independence is a key public service value, as well as a differentiator of public service media from its private-sector competitors, it is likely most public service media will retain a significant element of curation.

The requirement for universality also creates a strong incentive to ensure that there is a substantial foundation of shared information to which everyone in society should be exposed.

Recommendation systems in the BBC

The BBC is significantly larger in staff, output and audience than other European public service media organisations. It has a substantial research and development department and has been exploring the use of recommendation systems across a range of initiatives since 2008.[54]

In 2017, the BBC Datalab was established with the aim of helping audiences discover relevant content by bringing together data from across the BBC, augmented machine learning and editorial expertise.[55] It was envisioned as a central capability across the whole of the BBC (TV, radio, news and web) which would build a data platform for other BBC teams that would create consistent and relevant experiences for audiences across different products. In practice, this has meant collaborating with different product teams to develop recommendation systems.

---

54   See Annex 2 for more details.

55   BBC. (2019). 'Join the DataLab team at the BBC!'. *BBC Careers*. Available at: https://careerssearch.bbc.co.uk/jobs/job/Join-the-DataLab-team-at-the-BBC/40012; BBC Datalab. 'Machine learning at the BBC'. Available at: https://datalab.rocks/

The BBC now uses several recommendation systems, at different degrees of maturity, across different forms of media, including:

- **written content**, e.g. the BBC News app and some international news services, such as the Spanish-language BBC Mundo, recommending additional new stories[56]
- **audio-on-demand**, e.g. BBC Sounds recommending radio programmes and music mixes a user might like
- **short-form video**, e.g. BBC Sport and BBC+ (now discontinued) recommending videos the user might like
- **long-form video**, e.g. BBC iPlayer recommending TV shows or films the user might like.

Approaches to the development of recommendation systems

Public service media organisations have the choice to buy an external 'off the shelf' recommendation system or build it themselves.

The BBC initially used third-party providers of recommendation systems but, as part of a wider review of online services, began to test the pros and cons of bringing this function in-house. Building on years of their own R&D work, the BBC found they were able to build a recommendation system that not only matched but could outperform the bought-in systems. Once it was clear that personalisation would be central to the future strategy of the BBC, they decided to bring all systems in-house with the aim of being 'in control of their destiny'.[57] The perceived benefits include building up technical capability and understanding within the organisation, better control and integration of editorial teams, better alignment with public service values and greater opportunity to experiment in the future.[58]

The BBC has far greater budgets and expertise than most other public service media organisations to experiment with and develop recommendation systems. But many other organisations have also chosen to build their own products. Dutch broadcaster NPO has a small team of only four or five data scientists, focused on building 'smart but simple'

56   McGovern, A. (2019). 'Understanding public service curation: What do "good" recommendations look like?'. *BBC*.
     Available at: https://www.bbc.co.uk/blogs/internet/entries/887fd87e-1da7-45f3-9dc7-ce5956b790d2
57   Interview with Andrew McParland, Principal Engineer, BBC R&D (2021).
58   Commercial (i.e. non public service) BBC services however still use external recommendation providers. See: Taboola. (2021). 'BBC
     Global News Chooses Taboola as its Exclusive Content Recommendations Provider'. Available at:
     https://www.taboola.com/press-release/bbc-global-news-chooses-taboola-as-its-exclusive-content-recommendations-provider

recommendations in-house, having found third-party products did not cater to their needs. It is also important to them that they should be able to safeguard their audience data and be able to offer transparency to public stakeholders about the way their algorithms work, neither of which they felt confident about when using commercial providers.[59]

Several public service media organisations have joined forces through the EBU to develop PEACH[60] – a personalisation system that can be adopted by individual organisations and adapted to their needs. The aim is to share technical expertise and capacity across the public service media ecosystem, enabling those without their own in-house development teams to still adopt recommendation systems and other data-driven approaches. Although some public service media feel this is still not sufficiently tailored to their work,[61] others find it not only caters to their needs but that it embodies their public service mission through its collaborative approach.[62]

Although we are aware that some public service media continue to use third-party systems, we did not manage to secure research interviews with any organisations that currently do so.

How are public service media recommendation systems currently governed and overseen?

The governance of recommendation systems in public service media is created through a combination of data protection legislation, media regulation and internal guidelines. In this section, we outline the present and future regulatory environment in the UK and EU, and how internal guidelines influence development in the BBC and other public service media. Some public service media have reinterpreted their existing guidelines for operationalising public service values to make them relevant to the use of recommendation systems.

59   Interview with Arno van Rijswijk, Head of Data & Personalization, and Sarah van der Land, Digital Innovation Advisor, Nederlandse Publieke Omroep (NPO) (2021).

60   European Broadcasting Union. *PEACH*. Available at: https://peach.ebu.io/

61   Interview with Arno van Rijswijk, Head of Data & Personalization, and Sarah van der Land, Digital Innovation Advisor, Nederlandse Publieke Omroep (NPO) (2021).

62   Interview with Matthias Thar, Bayerische Rundfunk (2021).

The use of recommendation systems in public service media is not governed by any single piece of legislation or governance. Oversight is generated through a combination of the statutory governance of public service media, general data protection legislation and internal frameworks and mechanisms. This complex and fragmented picture makes it difficult to assess the effectiveness of current governance arrangements.

External regulation

The structures that have been established to regulate public service media are based around analogue broadcast technologies. Many are ill-equipped to provide oversight of public service media's digital platforms in general, let alone to specifically oversee the use of recommendation systems.

For instance, although Ofcom regulates all UK broadcasters, including the particular duties of public service media, its remit only covers the BBC's online platforms and not, for example, the ITV Hub or All 4. Its approach to the oversight of BBC iPlayer is to set broad obligations rather than specific requirements and it does not inspect the use of recommendation systems. Both the incentives and sanctions available to Ofcom are based around access to the broadcasting spectrum and so are not relevant to the digital dissemination of content. In practice this means that the use of recommendation systems within public service media are not subject to scrutiny by the communications regulator.

However, like all other organisations that process data, public service media within the European Union are required to comply with the General Data Protection Regulation (GDPR). The UK adopted this legislation before leaving the EU, though a draft Data Protection and Digital Information Bill ('Data Reform Bill') introduced in July 2022 includes a number of important changes, including removing the prohibition on automated decision-making, and maintaining restrictions for automated decision-making only if special categories of data are involved. The draft bill also introduces a new ground to allow the processing of

special categories of data for the purpose of monitoring and correcting algorithmic bias in AI systems. A separate set of provisions centred around fairness and explainability for AI systems is also expected as part of the Government's upcoming white paper on AI governance.

The UK GDPR shapes the development and implementation of recommendation systems because it requires:

- **Consent:** the UK GDPR requires that the use of personal data be made with freely-given, genuine and unambiguous consent from an individual. There are other lawful bases for processing personal data that do not require consent, including legal obligations, processing in a vital interest and processing for a 'legitimate interest' (a justification that public authorities cannot rely on if they are processing for their tasks as a public authority).

- **Data minimisation:** under Article 5(1), the 'data minimisation' principle of the UK GDPR states that personal data should be 'adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed'. Under Article 17 of the UK GDPR, the 'right to erasure' grants individuals the right to have personal data erased that is not necessary for the purposes of processing.

- **Automated decision-making, the right to be informed and explainability:** under the UK GDPR, data subjects have a right not to be subject to solely automated decisions that do not involve human intervention, such as profiling.[63] Where such automated decision-making occurs, meaningful information about the logic involved, the significance and the envisaged consequences of such processing need to be provided to the data subject (Article 15 (1) h). Separate guidance from the Information Commissioner's Office also touches on making AI systems explainable for users.[64]

Our interviews with practitioners indicated that GDPR compliance is foundational to their approach to recommendation systems, and that careful consideration must be paid to how personal data is collected

---

63   The Article 29 Working Group defines profiling in this instance as 'automated processing of data to analyze or to make predictions about individuals'.

64   Information Commissioner's Office and The Alan Turing Institute. (2021). *Explaining decisions made with AI.* Available at: https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/

and used. While the forthcoming Data Reform Bill makes several changes to the UK GDPR, most of these effects on the development and implementation of recommendation systems will likely continue under the current bill's language.

GDPR regulates the use of data that a recommendation system draws on, but there is not currently any legislation that specifically regulates the ways in which recommendation systems are designed to operate on that data, although there are a number of proposals in train at national and European levels.

In July 2022, the European Parliament adopted the Digital Services Act, which includes (in Article 24a) an obligation for all online platforms to explain, in their terms and conditions, the main parameters of their recommendation system and the options for users to modify or influence those parameters. There are additional requirements imposed on very large online platforms (VLOPs) to provide at least one option for each of their recommendation systems which is not based on profiling (Article 29). There are also further obligations for VLOPs in Article 26 to perform systemic risk assessments, including taking into account the design of the recommendation systems (Article 26 (2) a) and to implement steps to mitigate risk by testing and adapting their recommendation systems (Article 27 (1) ca).

In order to ensure compliance with the transparency provisions in the regulation, the Digital Services Act includes a provision that enables independent auditors and vetted researchers to have access to the data that led to the company's risk assessment conclusions and mitigation decisions (Article 31). This provision ensures oversight over the self-assessment (and over the independent audit) that companies are required to carry out, as well as scrutiny over the choices large companies make around their recommendation systems.

The draft AI Act proposed by the European Commission in 2021 also includes recommendation systems in its remit. The proposed rules require harm mitigations such as risk registers, data governance and human oversight but only make obligations mandatory for AI systems used in 'high-risk' applications. Public service media are not mentioned within this category, although due to their democratic significance it's possible they might come into consideration. Outside the high-risk categories, voluntary adoption is encouraged. These proposals are still at an early stage of development and negotiation and are unlikely to be adopted until at least 2023.

In another move, in January 2022 the European Commission launched a public consultation on a proposed European Media Freedom Act that aims to further increase the 'transparency, independence and accountability of actions affecting media markets, freedom and pluralism within the EU'. The initiative is a response to populist governments, particularly in Poland and Hungary attempting to control media outlets, as well as an attempt to bring media regulation up to speed with digital technologies. The proposals aim to secure 'conditions for [media markets'] healthy functioning (e.g. exposure of the public to a plurality of views, media innovation in the EU market)'. Though there is little detail so far, this framing could allow for the regulation of recommendation systems within media organisations.

In the UK, public service media are excluded from the draft Online Safety Bill which imposes responsibilities on platforms to safeguard users from harm. Ofcom, as well as the Digital Culture Media and Sport Select Committee, have called for urgent reform to regulation that would update the governance of public service media for the digital age. As of this report, there has been no sign of progress on a proposed Media Bill that would provide this guidance.

Internal oversight

Public service media have well-established practices for operationalising their mission and values through the editorial guidelines described earlier. But the introduction of recommendation systems has led many of them to reappraise these and, in some cases, introduce additional frameworks to translate these values for the new context.

The BBC has brought together teams from across the organisation to discuss and develop a set of machine learning engine principles, which they believe will uphold the Corporation's mission and values:[65]

- *Reflecting the BBC's values of trust, diversity, quality, value for money and creativity.*
- *Using machine learning to improve our audience's experience of the BBC*

---

65   Macgregor, M. (2021). *Responsible AI at the BBC: Our Machine Learning Engine Principles*. BBC Research and Development.. Available at: https://www.bbc.co.uk/rd/publications/responsible-ai-at-the-bbc-our-machine-learning-engine-principles

- *Carrying out regular review, ensuring data is handled securely and that algorithms serve our audiences equally and fairly*
- *Incorporating the BBC's editorial values and seeking to broaden, rather than narrow horizons.*
- *Continued innovation and human-in-the-loop oversight.*

These have then been adopted into a checklist for teams to use in practice:

> *'The MLEP [Machine Learning Engine Principles] Checklist sections are designed to correspond to each stage of developing a ML project, and contain prompts which are specific and actionable. Not every question in the checklist will be relevant to every project, and teams can answer in as much detail as they think appropriate. We ask teams to agree and keep a record of the final checklist; this self-audit approach is intended to empower practitioners, prompting reflection and appropriate action.*[66]

Reflecting on putting this into practice, BBC staff members observed that 'the MLEP approach is having real impact in bringing on board stakeholders from across the organisation, helping teams anticipate and tackle issues around transparency, diversity, and privacy in ML systems early in the development cycle'.[67]

Other public service media organisations have developed similar frameworks. Bayerische Rundfunk, the public broadcaster for Bavaria in Germany, found that their existing values needed to be translated into practical guidelines for working with algorithmic systems and developed ten core principles.[68] These align in many ways to the BBC principles but have additional elements, including a commitment to transparency and discourse, 'strengthening open debate on the future role of public service media in a data society', support for the regional innovation economy, engagement in collaboration and building diverse and skilled teams.[69]

66   Macgregor, M. (2021).

67   Boididou, C., Sheng, D., Moss, M. and Piscopo, A. (2021), 'Building Public Service Recommenders: Logbook of a Journey'. *RecSys '21: Proceedings of the 15th ACM Conference on Recommender Systems*, pp. 538–540. Available at: https://doi.org/10.1145/3460231.3474614

68   Bedford-Strohm, J., Köppen, U. and Schneider, C. (2020). 'Our AI Ethics Guidelines'. *Bayerisch Rundfunk*. https://www.br.de/extra/ai-automation-lab-english/ai-ethics100.html

69   Bedford-Strohm, J., Köppen, U. and Schneider, C. (2020).

In the Netherlands, public service broadcaster NPO along with commercial media groups and the Netherlands Institute for Sound and Vision drew up a declaration of intent.[70] Drawing on the European Union high-level expert group principles on ethics in AI, the declaration is a commitment to the responsible use of AI in the media sector. NPO are developing this into a 'data promise' that offers transparency to audiences about their practices.

Other stakeholders

Beyond these formal structures, the use of recommendation systems in public service media is shaped by these organisations' accountability to, and scrutiny by wider society.

All the public service media organisations we interviewed welcomed this scrutiny in principle and were committed to openness and transparency. Most publish regular blogposts about their work, present at academic conferences and invite feedback about their work. These, however, reach a small and specialist audience.

> There are limited opportunities for the broader public to understand and influence the use of recommendation systems. In practice, there is little accessible information about recommendation systems on most public service media platforms and even where it exists, teams admit that it is rarely read.

The Voice of the Listener and Viewer, a civil society group that represents audience interests in the UK, has raised concerns with the BBC about a lack of transparency in its approach to personalisation but has been dissatisfied with the response. The Media Reform Coalition has proposed that recommendations systems used in UK public service media should be co-designed with citizens' media assemblies and that

---

70   Media perspectives. (2021). 'Intentieverklaring voor verantwoord gebruik van KI in de media. [Letter of intent for responsible use of AI in the media]'. Available at: https://mediaperspectives.nl/intentieverklaring/

the underlying algorithms should be made public.[71]

Despite this low level of public engagement, public service media organisations were sensitive to external perceptions of their use of recommendation systems. Teams expected that, as public service media, they would be held to a higher standard than their commercial competitors. At the BBC in particular, staff frequently mentioned concerns about how their work might be seen by the press, the majority of which tends to take an anti-BBC stance. In practice, we have found little coverage of the BBC's use of algorithms outside of specialist publications such as *Wired*.

Public service media have a dual role, both as innovators in the use of recommendation services and as scrutineers of the impacts of new technologies. The BBC believes it has a 'critical contribution, as part of a mixed AI ecosystem, to the development of beneficial AI both technically, through the development of AI services, and editorially, by encouraging informed and balanced debate'.[72] At Bayerische Rundfunk, this combined responsibility has been operationalised by integrating the product team and data investigations team into an AI and Automation Lab. However, we are not aware of any instances where public service media have reported on their own products and subjected them to critical scrutiny.

Why this matters

The history of public service media, their current challenges and the systems for their governance are the framing context in which these organisations are developing and deploying recommendation systems. As with any technology, organisations must consider how the tool can be used in ways that are consistent with their values and culture and whether it can address the problems they face.

In his inaugural speech, BBC Director-General Tim Davie identified increased personalisation as a pillar of addressing the future role of

---

71   Grayson, D. (2021). *Manifesto for a People's Media*. Media Reform Coalition. Available at:
     https://drive.google.com/file/u/1/d/1_6GeXiDR3DGh1sYjFl_hbgV9HfLWzhPi/view?usp=embed_facebook

72   BBC. (2017). *Written evidence to the House of Lords Select Committee on Artificial Intelligence*. Available at:
     https://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-
     intelligence/written/70493.html

public service media in a digital world:[73]

> *'We will need to be cutting edge in our use of technology to join up the BBC, improving search, recommendations and access. And we must use the data we hold to create a closer relationship with those we serve. All this will drive love for the BBC as a whole and help make us an indispensable part of everyday life. And create a customer experience that delivers maximum value.'*

But recommendation systems also crystallise the current existential dilemmas of public service media. The development of a technology whose aim is optimisation requires an organisation to be explicit about what and who it is optimising for. A data-driven system requires an institution to quantify those objectives and evaluate whether or not the tool is helping them to achieve them.

This can seem relatively straightforward when setting up a recommendation system for e-commerce, for example, where the goal is to sell more units. Other media organisations may also have clear metrics around time spent on a platform, advertising revenues or subscription renewals.

In this instance, the broadly framed public service values that have proven flexible to changing contexts in the past are a hindrance rather than a help. A concept like 'diversity' is hard to pin down and feed into a system.[74] Organisations that are supposed to serve the public as both citizens and consumers must decide which role gets more weight.

Recommendation systems might offer an apparently obvious solution to the problem of falling public service media audience share – if you are able to better match the vast amount of content in public service media catalogues to listeners and viewers, you should be able to hold and grow your audience. But is universality achieved if you reach more people but they don't share a common experience of a service? And how do you measure diversity and ensure personalised recommendations still offer a balance of content?

73   BBC Media Centre. (2020). *Tim Davie's introductory speech as BBC Director-General.* Available at:
     https://www.bbc.co.uk/mediacentre/speeches/2020/tim-davie-intro-speech
74   Hildén, J. (2021). 'The Public Service Approach to Recommender Systems: Filtering to Cultivate'. *Television & New Media*, 23(7).
     Available at: https://doi.org/10.1177/15274764211020106

'The introduction of algorithmic systems will force [public service media] to express its values and goals as measurable key performance indicators, which could be useful and perhaps even necessary. But this could also create existential threats to the institution by undermining the core principles and values that are essential for legitimacy.'[75]

Recommendation systems force product teams within public service media organisations to settle on an interpretation of public service values, at a time when the regulatory, social and political context makes them particularly unclear.

It also means that this interpretation will be both instantiated and then systematised in a way that has never previously occurred. As we saw with the example of the impartiality guidelines of the BBC, individuals and teams have historically made decisions under a broad governance framework and founded on editorial judgement. Inconsistencies in those judgements could be ironed out through the multiplicity of individual decisions, the diversity of contexts and the number of different decision-makers. Questions of balance could be considered over a wider period of time and breadth of output. Evolving societal norms could be adopted as audience expectations change.

However, building a decision-making system sets a standardised response to a set of questions and repeats that every time. In this way it nails an organisation's colours to one particular mast and then replicates that approach repeatedly.

---

75   Sørensen, J.K. and Hutchinson, J. (2018). 'Algorithms and Public Service Media'. *Public Service Media in the Networked Society: RIPE@2017*, pp.91–106. Available at:
http://www.nordicom.gu.se/sites/default/files/publikationer-hela-pdf/public_service_media_in_the_networked_society_ripe_2017.pdf

Stated goals and potential risks of using recommendation systems in public service media

Organisations deploy recommendation systems to address certain objectives. However, these systems also bring potential risks. In this chapter, we look at what public service media aim to achieve through deploying recommendation systems and the potential drawbacks.

Stated goals of recommendation systems

In this section, we look at the stated objectives for the use of recommendation systems and the degree to which public service media reference those objectives and motivations when justifying their own use of recommendation systems.

Recommendation systems bring several benefits to different actors, including users who access the recommendations (in the case of public service media, audiences), as well as the organisations and businesses that maintain the platforms on which recommendation systems operate. Some of the effects of recommendation systems are also of broader societal interest, especially where the recommendations interact with large numbers of users, with the potential to influence their behaviour. Because they serve the interests of multiple stakeholders,[76] recommendation systems support data-based value creation in multiple ways, which can pull in different directions.[77]

---

76   Milano, S., Taddeo, M. and Floridi, L. (2021). 'Ethical aspects of multi-stakeholder recommendation systems'. *The Information Society*, 37(1). Available at: https://doi.org/10.1080/01972243.2020.1832636; Abdollahpouri, H., Adomavicius, G., Burke, R. et al. (2020). 'Multistakeholder recommendation: Survey and research directions'. *User Modeling and User-Adapted Interaction*, pp.127–158. Available at: https://doi.org/10.1007/s11257-019-09256-1

77   Tempini, N. (2017). 'Till data do us part: Understanding data-based value creation in data-intensive infrastructures'. *Information and Organization*, 27(4). Available at: http://dx.doi.org/10.1016/j.infoandorg.2017.08.001

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain… and recommend?   52

Four key areas of value creation are:

1. **Reducing information overload for the receivers of recommendations:** It would be overwhelming for individuals to trawl the entire catalogue of Netflix or Spotify, for example. Their recommendation systems reduce the amount of content to a manageable number of choices for the audience. This creates value for users.

2. **Improved discoverability of items:** E-commerce sites can recommend items they are particularly keen to sell, or direct people to niche products for which there is a specific customer base. This creates value for businesses and other actors that provide the items in the recommender's catalogue. It can also be a source of societal value, for example where improved discoverability increases the diversity of news items that are accessed by the audience.

3. **Attention capture:** Targeted recommendations which cater to users' preferences encourage people to spend more time on services, generating revenue through subscriptions or advertising. This is a source of economic value for platform providers, who monetise attention via advertising revenue or paid subscriptions. But it can also be a source of societal value, if it means that people pay more attention to content that has public service value, in line with the mandate for universality.

4. **Data gathering to derive business insights and analysis:** For example, platforms gain valuable insights into their audience through A/B testing which enables them to plan marketing campaigns or commission content. This is a source of economic value, when it is used to derive business insights. But under appropriate conditions, it could be a source of societal value, for example by enabling socially responsible scientific research (see our recommendations below).

We explored how these objectives map to the motivations articulated by public service media organisations for their use of recommendation systems.

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?   53

1. Reducing information overload

'Under conditions of information abundance and attention scarcity, the modern challenges to the realisation of media diversity as a policy goal lie less and less in guaranteeing a diversity of supply and more in the quest to create the conditions under which users can actually find and choose between diverse content.'[78]

We heard from David Graus: 'So finding different ways to enable users to find content is core there. And in that context, I think recommender systems really serve to be able to surface content that users may not have found otherwise, or may surface content that users may not know they're interested in'.

Being able to access more relevant content from large catalogues is a source of value for users of a recommendation system. Without such systems, platforms that host large catalogues would not be viable. For example, a large e-commerce catalogue would be of little use to consumers without the ability to navigate it efficiently, identifying the items that might correspond to their needs. Navigating the content uploaded on social media would be impossible, and not meaningful, without something to structure and organise the experience. By automating the curation process, recommendation systems provide value to users, enabling them to have a curated experience of the item catalogue available.

2. Improved discoverability

Public service media also deploy recommendation systems with the objective of showcasing much more of their vast libraries of content. BBC Sounds, for example, has more than 200,000 items available, of which only a tiny amount can be surfaced either through broadcast schedules or an editorially curated platform. Recommendation systems can potentially unlock the long tail of rarely viewed content and allow individuals' specific interests to be met.

---

78   Helberger, N., Karppinen, K. and D'Acunto, L. (2018). 'Exposure diversity as a design principle for recommender systems'. *Information, Communication & Society*, 21(2). Available at: https://doi.org/10.1080/1369118X.2016.1271900

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

54

They can also, in the view of some organisations, meet the public service obligation of diversity by exposing audiences to a greater variety of content. Recommendation systems need not simply cater to, or replicate people's existing interests but can actively push new and surprising content.[79]

This approach is also deployed in commercial settings, notably in Spotify's 'Discover' playlists, as novelty is also required for audience retention. Additionally, some public service media organisations, such as Swedish Radio and NPO, are experimenting with approaches that promote content they consider particularly high in public value.

Traditional broadcasting provides one-to-many communication. Through personalisation, platforms have created a new model of many-to-many communication, creating 'fragmented user needs'.[80] Public service media must now grapple with how they create their own way of engaging in this landscape. The BBC's ambition for the iPlayer is to make output, 'accessible to the audience wherever they are, whatever devices they are using, finding them at the right moments with the right content'.[81]

Jonas Schlatterbeck, ARD (German public broadcaster), takes a similar view:

> 'We can't actually serve majorities anymore with one content. It's not like the one Saturday night show that will attract like half of the German population [...] but more like tiny mosaic pieces of different content that are always available to pretty much everyone but that are actually more targeted.'[82]

3. Attention capture

The need to maintain audience reach in a fiercely competitive digital landscape was mentioned by almost every public service media

79   Interview with David Graus, Lead Data Scientist, Randstad Groep Nederland (2021). This point was also captured in separate studies of public service media organisations – see: Hildén, J. (2021). 'The Public Service Approach to Recommender Systems: Filtering to Cultivate'. *Television & New Media*, 23(7). Available at: https://doi.org/10.1177/15274764211020106

80   Interview with Uli Köppen, Head of AI + Automation Lab, Co-Lead BR Data, Bayerische Rundfunk (2021).

81   BBC. (2021). *BBC Annual Plan 2021-22*. Available at:
http://downloads.bbc.co.uk/aboutthebbc/reports/annualplan/annual-plan-2021-22.pdf

82   Interview with Jonas Schlatterbeck, Head of Content ARD Online & Leiter Programmplanung, ARD (2021).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

55

organisation we spoke to. Universality, the obligation to reach every section of society, is central to the public service remit.

And if public service media lose their audience to their digital competitors, they cannot deliver the other societal benefits within their mission. As Koen Muylaert of Belgian VRT said: 'we want to inspire people, but we also know that you can only inspire people if they intensively use your products, so our goal is to increase the activity on our platform as well. Because we have to fight for market share'.[83]

The assumption among most public service media organisations is that recommendation systems improve engagement, although there is still little conclusive evidence of this in academic literature. The BBC has specific targets for 16-34 year-olds to use the iPlayer and BBC Sounds, and staff consider recommendations as a route to achieving those metrics.[84]

From our interview with David Caswell, Executive Product Manager, BBC News Labs:

> 'We have seen that finding in our research on several occasions that there's sort of some transition that audiences and particularly younger audiences have gone through where there's an expectation of personalization they don't expect to be doing the same thing again and again and again, and in terms of active searching for things they expect they expect a personalized experience... There isn't a lot of tolerance, increasingly with younger and digitally native audiences for friction in the experience. And so personalization is a major technique for removing friction from the experience because audience members don't have to do all the work of discovery and selection and so on, they can have that done for them that this is.'[85]

Across the teams we interviewed from European public service media organisations there was widespread consensus that audiences now expect content to be personalised. Netflix and Spotify's use

83   Interview with Koen Muylaert, Project Lead, VRT data platform and data science initiative, Vlaamse Radio-
     en Televisieomroeporganisatie (VRT) (2021).

84   BBC. (2021). *BBC Annual Plan 2021-22*. Available at:
     http://downloads.bbc.co.uk/aboutthebbc/reports/annualplan/annual-plan-2021-22.pdf

85   Interview with David Caswell, Executive Product Manager, BBC News Labs (2021).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

56

of recommendation systems was described as a 'gold standard'
for public service media organisations to aspire to. But few of our
interviewees offered evidence to support this view of audience
expectations.

> 'I see the risk that when we are compared with some of our competitors
> that are dabbling with a much more sophisticated personalisation,
> there is a big risk of our services being perceived as not adaptable and
> not relevant enough.'[86]

4. Data gathering and behavioural interventions

Recommendation systems collect and analyse a wealth of data in
order to serve personalised recommendations to their users. The data
collected often pertains to user interactions with the system, including
data that is produced as a result of interventions on the part of the
system that are intended to influence user behaviour (interventional
data).[87] For example, user data collected by a recommendation system
may include data about how different users responded to A/B tests,
so that the system developers can track the effectiveness of different
designs or recommendation strategies in stimulating some desired user
behaviour.

Interventional data can thus be used to support targeted behavioural
interventions, as well as scientific research into the mechanisms
that underpin the effectiveness of recommendations. This marks
recommendation systems as a key instrument of what Shoshana
Zuboff has called a system of 'surveillance capitalism'.[88] In this system,
platforms extract economic value from personal data, usually in the form
of advertising revenue or subscriptions, at the expense of the individual
autonomy afforded to individual users of the technology.

---

86   Interview with Olle Zachrison, Deputy News Commissioner & Head of Digital News Strategy, Swedish Radio (2021).

87   Greene, T., Martens, D. and Shmueli, G. (2022) 'Barriers to academic data science research in the new realm of algorithmic behaviour
        modification by digital platforms'. *Nature Machine Intelligence*, 4(4), pp. 323–330. Available at:
        https://doi.org/10.1038/s42256-022-00475-7

88   Zuboff, S. (2015). 'Big other: Surveillance Capitalism and the Prospects of an Information Civilization'. *Journal of Information
        Technology*, 30(1). Available at: https://doi.org/10.1057/jit.2015.5

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

57

> As access to the services provided by the platforms becomes essential to daily life, users increasingly find themselves tracked in all aspects of their online experience, without meaningful options to avoid it. The possibility of surveillance constitutes a grave risk associated with the use of recommendation systems.

Because recommendation systems have been mainly researched and developed in commercial settings, many of the techniques and types of data collected work within this logic of surveillance.[89] However, it is also possible to envisage uses of recommendation systems that do not obey the same logic.[90] Recommendation systems used by public service media are a case in point. Public service media organisations are in a position to decide which data to collect and use in the service of creating public value, scientific value and individual value for their audiences, instead of economic value that would be captured by shareholders.[91]

Examples of public value that could be created from user data include insights into effective and impartial communication that serves the public interest and fosters community building. Social science research into the effectiveness of behavioural interventions, and basic research into the psychological mechanisms that underpin audience's trust in recommendations would contribute to the creation of scientific value from behavioural data. From the perspective of the audience, value could be created by fostering user empowerment to learn more about their own interests and develop their tastes, letting users feel more in control and understand the value of the content that they can access.

We found little evidence of public service media deploying recommendation systems with the explicit aim of capturing data on their audiences and content or deriving greater insights. On the contrary, interviewees stressed the importance of data minimisation and privacy. At Bayerische Rundfunk

89   van Dijck, J. (2014). 'Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology'. *Surveillance & Society*, 12(2). Available at: https://doi.org/10.24908/ss.v12i2.4776; Srnicek, N. (2017). *Platform capitalism*. Polity.

90   Lane, J. (2020). *Democratizing Our Data: A Manifesto*. MIT Press.

91   Tempini, N. (2017). 'Till data do us part: Understanding data-based value creation in data-intensive infrastructures'. *Information and Organization,* 27(4). Available at: http://dx.doi.org/10.1016/j.infoandorg.2017.08.001

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?   58

for example, a product owner said that the collection of demographic data on the audience was a red line that they would not cross.[92]

However, we did find that most public service media organisations introduced recommendation systems as part of a wider deployment of automated and data-driven approaches. In many cases, these are accompanied by significant organisational restructures to create new ways of working adapted to the technologies, as well as to respond to the budget cuts that almost all public service media are facing.

Public service media organisations are often fragmented, with teams separated by region and subject matter and with different systems for different channels and media that have evolved over time. The use of recommendation systems requires a consistent set of information about each item of content (commonly known as metadata). As a result, some public service media have started to better connect different services so that recommendation systems can draw on them.

For instance, Swedish Radio has overhauled its entire news output to improve its digital service, creating standalone items of content that do not need to be slotted into a particular programme or schedule but can be presented in a variety of contexts. Alongside this, it has introduced a scoring system to rank its content against its own public values, prompting a rearticulation of those values as well as a renewed emphasis on their importance.

Bayerische Rundfunk (BR) is creating a new infrastructure for the consistent use of data as a foundation for the future use of recommendation systems. This is already allowing for news stories to automatically upload data specific to different localities, as well as generating automated text on data-heavy stories such as sports results. This allows BR to cover a broader range of sports and cater to more specialist interests, as well as freeing up editorial teams from mundane tasks.

While there is not a direct objective of behavioural intervention and data capture at present, the introduction of recommendation systems is part of a wider orientation towards data-driven practices across public

92   Interview with Matthias Thar, Bayerische Rundfunk (2021).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

59

service media organisations. This has the potential to enable wider data collection and analysis to generate business insights in the future.

Conclusion

We find that public service media organisations articulate similar objectives to the field more broadly, in their motivations for deploying recommendation systems, although unlike commercial actors, they do not currently use recommendations for the explicit aim of data capture and behavioural intervention. In some respects they reframe these established motivations to align with their public service mission and values.

Many staff across public service media organisations display a belief that because the organisation is motivated by public service values, and produces content that adheres to those values, the use of recommendation systems to filter that content is a furtherance of their mission.

This has meant that staff at public service media organisations have not always critically examined whether the recommendation system itself is operating in accordance with public service values.

However, public service media organisations have begun to put in place principles and governance mechanisms to encourage staff to explicitly and systematically consider how the development of their systems furthers their public service values. For example, the BBC published its Machine Learning Engine Principles in 2019 and subsequently continues to iterate on a checklist for project teams to put those principles into practice.[93]

Public service media organisations are also in the early stages of developing new metrics and methods to measure the public service value of the outputs of the recommendation systems, both with explicit measures of 'public service value' and implicitly through evaluation by editorial staff. We explore these more in our chapter on evaluation and in our case studies on the BBC's use of recommendation systems.

93   Macgregor, M. (2021). *Responsible AI at the BBC: Our Machine Learning Engine Principles*. BBC Research and Development. Available
     at: https://www.bbc.co.uk/rd/publications/responsible-ai-at-the-bbc-our-machine-learning-engine-principles

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain… and recommend?

60

Additionally, we found that alongside these stated motivations, public service media interviewees had internalised a set of normative values around recommendation systems. When asked to define what a recommendation system is in their own terms, they spoke of systems helping users to find 'relevant', 'useful', 'suitable', 'valuable' or 'good' content.[94]

This framing around user benefit obscures the fact that the systems are ultimately deployed to achieve organisations' goals, and so if they are 'relevant' or 'useful' this is because that helps achieve the organisations' goals, not because of an inherent property of the system.[95] It also adopts the vocabulary of commercial recommendation systems (e.g. targeted advertising options encourage users to opt for more 'relevant' adverts) which the Competition and Markets Authority has identified as problematic. This indicates that public service media are essentially adopting the paradigm established by the use of commercial recommendation systems.

Potential risks from recommendation systems

In this section, we explore some of the ethical risks associated with the use of recommendation systems and how they might manifest in uses by public service media.

A review of the literature on recommendation systems helps identify some of the potential ethical and societal risks that have been raised in relation to their use beyond the specific context of public service media. Milano et al highlight six areas of concern for recommendation systems in general:[96]

1.   **Privacy risks** to users of a recommendation system: including direct risks from non-compliance with existing privacy regulations and/or malicious use of personal data, and indirect risks resulting from data

---

94   This is not unique to the BBC, and many academic papers and industry publications also reflect a similar implicit normative framework in their definitions of recommendation systems.

95   The organisations' goals are not necessarily in tension with that of the users, e.g. helping audiences finding more relevant content might help audiences get better value for money (which is a goal of many public service media organisations) but that is still goal which shapes how the recommendation system is developed, rather than a necessary feature of the system.

96   Milano, S., Taddeo, M. and Floridi, L. (2020). 'Recommender systems and their ethical challenges'. *AI & Society*, 35, pp.957–967. Available at: https://doi.org/10.1007/s00146-020-00950-y

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain… and recommend?   61

leaks, deanonymisation of public datasets or unwanted exposure of inferred sensitive characteristics to third parties.

2. **Problematic or inappropriate content** could be recommended and amplified by a recommendation system.

3. **Opacity** in the operation of a recommendation system could lead to limited accountability and lower the trustworthiness of the recommendations.

4. **Autonomy:** recommendations could limit users' autonomy by manipulating their beliefs or values, and by unduly restricting the range of meaningful options that are available to them.

5. **Fairness** constitutes a challenge for any algorithmic system that operates using human-generated data and is therefore liable to (re) produce social biases. Recommendation systems are no exception, and can exhibit unfair biases affecting a variety of stakeholders whose interests are tied to recommendations.

6. **Social externalities** such as polarisation, the formation of echo chambers, and epistemic fragmentation, can result from the operation of recommendation systems that optimise for poorly defined objectives.

How these risks are viewed and addressed by public service media

In this section, we examine the extent to which ethical risks of recommendation systems, identified in the literature, are present in the development and use of recommendation systems in practice by public service media.

1. Privacy

The data gathering and operation of recommendation systems can pose direct and indirect privacy risks. Direct privacy risks come from how personal data is handled by the platform, as its collection, usage and storage need to follow procedures to ensure prior consent from individual users. In the context of EU law, these stages are covered by General Data Protection Regulation (GDPR).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

62

Indirect privacy risks arise when recommendation systems expose sensitive user data unintentionally. For instance, indirect privacy risks may come about as a result of unauthorised data breaches, or when a system reveals sensitive inferred characteristics about a user (e.g. targeted advertising for baby products could indicate a user is pregnant).

Privacy relates to a number of public service values: **independence** (act in the interest of audiences), **excellence** (high standards of integrity) and **accountability** (good governance).

Privacy was raised as a potential risk by every interviewee from a public service organisation. Specifically, public service media were concerned about users' consent to the use of their data, emphasising data security as a key concern for the responsible collection and use of user data.[97] Several interviewees stressed that public service media organisations do not generally require mandatory sign-in for certain key products such as news. Other services, focusing more on entertainment, such as BBC iPlayer, do require sign-on, but the amount of personal data collected is limited.

Sebastien Noir, Head of Software, Technology and Innovation at the European Broadcasting Union, emphasised how the need to comply with privacy regulations in practice means that projects have to jump through several hoops with legal teams before trials with user data are allowed. While this uses up time and resources in project development, it also means that robust measures are in place to protect users from direct threats to privacy. Koen Muylaert, at Belgian VRT, also spoke to us about how there is a distinction between personal data, which poses privacy risks, and behavioural data, which may be safer to use for public service media recommendation systems and which they actively monitor.[98]

None of the organisations that we interviewed spoke to us about indirect threats to privacy or ways to mitigate them.

97   Interview with Jonas Schlatterbeck, Head of Content ARD Online & Leiter Programmplanung, ARD (2021).

98   Interview with Koen Muylaert, Project Lead, VRT data platform and data science initiative, Vlaamse Radioen Televisieomroeporganisatie (VRT) (2021).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

63

2. Problematic or inappropriate content

Open recommendation systems on commercial platforms that host limitless, user-generated content have a high risk of recommending low quality or harmful content. This risk is lower for public service media that deploy closed recommendation systems to filter their own catalogue of content which has already been extensively scrutinised for quality and adherence to editorial guidelines. Nonetheless, some risk may still exist for closed recommendation systems, such as the risk of recommended age-inappropriate content to younger users.

The risk of inappropriate content relates to the public service media values of **excellence** (high standards of integrity, professionalism and quality) and **independence** (completely impartial and independent from political commercial and other influences and ideologies).

In interviews, many members of public service media staff were generally confident that recommendations would be of high quality and represent public service values because the content pool had already passed that test. Nonetheless, some staff identified a risk that the system could surface inappropriate content, for example, archive items that include sexist or racist language that is no longer acceptable or through the juxtaposition of items that could be jarring.

However, a more commonly identified potential risk arises in connection to independence and impartiality. Many of the interviewees we spoke to mentioned that the algorithms used to generate user recommendations needed to be impartial. The BBC and other public service media organisations have traditionally operated a policy of 'balance over time and output', meaning a range of views on a subject or party political voices will be heard over a given period of programming on a specific channel. However, recommendation systems disrupt this. The audience is no longer exposed to a range of content broadcast through channels. Instead, individuals are served up specific items of content without the balancing context of other programming. In this way they may only encounter one side of an argument.

Therefore, some interviewees expressed that fine-tuning balanced recommendations are especially important in this context. This is an area where the close integration of editorial and technical teams was seen to be essential.

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

64

3. Opacity of the recommendation

Like many other algorithmic systems, many recommendation systems
operate as black boxes whose internal workings are sometimes
difficult to interpret, even for their developers. The process by which a
recommendation is generated is often not transparent to individual users
or other parties that interact with a recommendation system. This can
have negative effects, by limiting the accountability of the system itself,
and diminishing the trust that audiences put in the good operation of the
service.

Opacity is a challenge to the public service media values of
independence (autonomous in all aspects of the remit) and
accountability (be transparent and subject to constant public scrutiny).
The issue of opacity and the risks that it raises was touched upon in
several of our interviews.

The necessity to exert more control over the data and algorithms
used for building recommendation systems was among the
motivations for the BBC in bringing their development in house. The
same is true of other public service media in Europe. While most
European broadcasters did not choose to bring the development of
recommendation systems in house, many of them now rely on PEACH,
a recommendation system developed collaboratively by several public
service media organisations under the umbrella of the European
Broadcasting Union (EBU).

Previously, the BBC as well as other public service media had relied
on external commercial contractors to build the recommendation
systems they used. This however meant that they could exert little
control over the data and algorithms used, which represented a risk.
In the words of Sebastien Noir, Head of Software, Technology and
Innovation at the EBU:

> 'As a broadcaster, you are defined by what you
> promote to the people, that's your editorial line.
> This is, in a way, also your brand or your user
> experience. If you delegate that to a third party
> company, [...] then you have a problem, because

you have given your very identity, the way you are perceived by the people to a third party company […] No black box should be your editorial line.'[99]

But bringing the development of recommendation systems in-house does not solve all the issues connected with the opacity of these systems. Jannick Sørenson, Associate Professor in Digital Media at Aalborg University, summarised the concern:

'I think the problem of the accountability, first within the public service institution, is that editors, they have no real chance to understand what data scientists are doing. And data scientists, neither they do. […] And so the dilemma here is that it requires a lot of specialised knowledge to understand what is going on inside this process of computing recommendation[s]. Right. And, I mean, with Machine Learning, it's become literally impossible to follow.'[100]

Sørenson highlighted how the issue of opacity arises both internally and externally for public service media.

Internally to the institution, the opacity of the systems utilised to produce recommendations hinders the collaboration of editorial and technical staff. Some public service media organisations, such as Swedish Radio, have tried to tackle this issue by explicitly having both a technical and an editorial project lead, while Bayerische Rundfunk have established an interdisciplinary team with their AI and Automation Lab. [101]

---

99 Interview with Sébastien Noir, Head of Software, Technology and Innovation, and Dmytro Petruk, Developer, European Broadcasting Union (2021).

100 Interview with Jannick Kirk Sørensen, Associate Professor in Digital Media, Aalborg University (2021).

101 We explore these examples in more detail later in the chapter.

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain… and recommend?

66

Documentation is another approach taken by public service media organisations to reduce the opacity of the system. For example, the BBC's Machine Learning Engine Principles checklist (as of version 2.0) explicitly asks teams to document what their model does and how it was created, e.g. via a data science decision log, and to create a Plain English explanation or visualisation of the model to communicate the model's purpose and operation.

Externally, public service media struggle to provide effective explanations to audiences about the systems that they use. The absence of industry standards for explanation and transparency was identified as a risk. Olle Zachrison, Deputy News Commissioner & Head of Digital News Strategy, Swedish Radio, also expressed this worry:

> 'One particular risk, I think, with all these kind of more automatic services, and especially with the introduction of […] AI powered services, is that the audience doesn't understand what we're doing. And […] I know that there's a big discussion going on at the moment, for example, about Explainable AI. How should we explain in a better way what the services are doing? […] I think that there's a very big need for kind of industry dialogue about setting standards here, you know.'[102]

Other interviewees, however, highlighted that the use of explanations has limited efficacy in addressing the external opacity of individual recommendations, since users rarely pay attention to them. Sarah van der Land, Digital Innovation Advisor at NPO in the Netherlands, cited internally conducted consumer studies as evidence that audiences might not care about explanations:

---

102   Interview with Olle Zachrison, Deputy News Commissioner & Head of Digital News Strategy, Swedish Radio (2021).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain… and recommend?

67

'Recently, we did some experiments also on data insight, into what extent our consumers want to have feedback on why they get a certain recommendation? And yeah, unfortunately, our research showed that a lot of consumers are not really interested in the why. […] Which was quite interesting for us, because we thought, yeah, of course, as a public value, we care about our consumers. We want to elaborate on why we do the things we do and why, based on which data, consumers get these recommendations. But yeah, they seem to be very little interested in that.'[103]

This finding indicates that pursuing this strategy has limited practical effects in improving the value of recommendations for audiences. David Graus, Lead Data Scientist, Randstad Groep Nederland, also told us that he is sceptical of the use of technical explanations, but that 'what is more important is for people to understand what a recommender system is, and what it aims to do, and not how technically a recommendation was generated.'[104] This could be achieved by providing high-level explanations of the processes and data that were used to produce the recommendations, instead of technical details of limited interest to non-technical stakeholders.

4. Autonomy

Research on recommendation systems has highlighted how they could pose risks to user autonomy, by restricting people's access to information and by potentially being used to shape preferences or emotions. Autonomy is a fundamental human value which 'generally can

---

103   Interview with Arno van Rijswijk, Head of Data & Personalization, and Sarah van der Land, Digital Innovation Advisor, Nederlandse Publieke Omroep (2021).

104   Interview with David Graus, Lead Data Scientist, Randstad Groep Nederland (2021).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

68

be taken to refer to a person's effective capacity for self-governance'.[105] Writing on the concept of human autonomy in the age of AI, Prunkl distinguishes two dimensions of autonomy: one *internal*, relating to the authenticity of the beliefs and values of a person; and the other *external*, referring to the person's ability to act, or the availability of meaningful options that enables them to express agency.

The risk to autonomy relates to the public service media value of **universality** (creating a public sphere, in which all citizens can form their own opinions and ideas, aiming for inclusion and social cohesion).

Public service media historically have made choices on behalf of their audiences in line with what the organisation has determined is in the public interest. In this sense audiences have limited autonomy due to public service media organisations restricting individuals' access to information, albeit with good intentions.

The use of recommendation systems could, in one respect, be seen as increasing the autonomy of audiences. A more personalised experience, that is more tailored to the individual and their interests, could support the 'internal' dimension of autonomy, because it could enable a recommendation system to more accurately reflect the beliefs and values of an individual user, based on what other users of that demographic, region or age might like.

At the same time, public service media strive to 'create a public sphere, in which all citizens can form their own opinions and ideas, aiming for inclusion and social cohesion'.[106] There is a risk in using recommendation systems that public service media might filter information in such a way that they inhibit people's autonomy to form their views independently.[107]

By design, recommendation systems tailor recommendations to a specific individual, often in such a way where these recommendations are not visible to other people. This means individual members of the audience may not share a common context or may be less aware of

105   Prunkl, C. (2022). 'Human autonomy in the age of artificial intelligence'. *Nature Machine Intelligence*, 4, pp.99–101. Available at: doi: https://doi.org/10.1038/s42256-022-00449-9

106   European Broadcasting Union. (2012). *Empowering Society: A Declaration on the Core Values of Public Service Media*, p. 4. Available at: https://www.ebu.ch/files/live/sites/ebu/files/Publications/EBU-Empowering-Society_EN.pdf;

107   Interview with David Caswell, Executive Product Manager, BBC News Labs (2021).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?   69

what information others have access to, a condition that Milano et al have called 'epistemic fragmentation'.[108] Coming to an informed opinion often requires being able to have meaningful conversations about a topic with other people. If recommendations isolate individuals from each other, then this may undermine the ability of audiences to form authentic beliefs and reason about their values. Since this ability is essential to having autonomy, epistemic fragmentation poses a risk.

Recommendations are also based on an assumption that there is such a thing as a single, legible individual for whom content can be personalised. In practice, people's needs vary according to context and relationships. They may want different types of content at different times of day, whether they are watching videos with family or listening to the news in the car, for example. However, contextual information is difficult to factor in a recommendation, and doing so requires access to more user data which could pose additional privacy risks. Moreover, recommendations are often delivered via a user's account with a service that uses recommendation systems. However, some people may choose to share accounts, create a joint one or maintain multiple personal accounts to compartmentalise different aspects of their information needs and public presence.[109]

Finally, the use of recommendation systems by public service media can pose a risk to autonomy when the categories that are used to profile users are not accurate, not transparent or not easily accessible and modifiable by the users themselves. This concern is linked to the opacity of the system, but it was not addressed explicitly as a risk to user autonomy in our interviews.

As above, several interviews highlighted that internal research indicates users do not want more explanations and control over the recommendation system, when this comes at the cost of a frictionless experience. If so, public service media need to consider whether there is a trade-off between supporting autonomy and the ease of use of a recommendation system, and research alternative strategies to provide audiences with more meaningful opportunities to participate in the construction of their digital profiles.

108 Milano, S., Mittelstadt, B., Wachter, S. and Russell, C. (2021), 'Epistemic fragmentation poses a threat to the governance of online targeting'. *Nature Machine Intelligence*. Available at: https://doi.org/10.1038/s42256-021-00358-3

109 Milano, S., Taddeo, M. and Floridi, L. (2021). 'Ethical aspects of multi-stakeholder recommendation systems'. *The Information Society*, 37(1). Available at: https://doi.org/10.1080/01972243.2020.1832636

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

70

5. Fairness

Researchers have documented how the use of machine learning and AI in applications ranging from credit scoring to facial recognition,[110] medical triage to parole decisions,[111] advert delivery[112] to automatic text generation[113] and many others, often leads to unfair outcomes which perpetuate historical social biases or introduce new, machine-generated ones. Given the pervasiveness of these systems in our societies, this has given rise to increasing pressure to improve their fairness, which has contributed to a burgeoning area of research.

This risk relates to the public service media value of **universality** (reach all segments of society, with no-one excluded) and **diversity** (support and seek to give voice to a plurality of competing views – from those with different backgrounds, histories and stories. Help build a more inclusive, less fragmented society).

Developers of algorithmic systems today can draw on a growing array of technical approaches to addressing fairness issues; however, fairness remains a challenging issue that cannot be fully solved by technical fixes. Instead, as Wachter et al argue in the context of EU law, the best approach may be to recognise that algorithmic systems are inherently and inevitably biased, and to put in place accountability mechanisms to ensure that there are no biases that perpetuate unfair discrimination, but to the contrary biases are used to help to redress historical injustices.[114]

Recommendation systems are no exception. Biases in recommendation can arise at a variety of levels and for different stakeholders. From the perspective of users, a recommendation system could be unfair if the quality of the recommendations varies across users. For example, if a music recommendation system is much worse at predicting the tastes of

110   Buolamwini, J. and Gebru, T. (2018). 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification'. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency. Conference on Fairness, Accountability and Transparency, PMLR*, pp. 77–91. Available at: https://proceedings.mlr.press/v81/buolamwini18a.html

111   Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016). 'Machine Bias'. *ProPublica*. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

112   Sweeney, L. (2013). 'Discrimination in online ad delivery'. *arXiv*. Available at: https://doi.org/10.48550/arXiv.1301.6822

113   Noble, S. U. (2018). *Algorithms of Oppression*. New York: New York University Press; Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. (2021). 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?'. *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp.610–623. Available at: https://doi.org/10.1145/3442188.3445922

114   Wachter, S., Mittelstadt, B. and Russell, C. (2020). 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI'. *Computer Law & Security Review*, 41. Available at: http://dx.doi.org/10.2139/ssrn.3547922

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

71

and serving interesting recommendations to a minority group, this could be unfair.

Recommendations could also be unfair from a provider perspective. For instance, one recent study found a film recommendation system trained on a well-known dataset (MovieLens 10M), and designed to optimise for relevance to users, systematically underrepresented films by female directors.[115] This example illustrates a phenomenon that is more pervasive. Since recommendation systems are primarily built to optimise for user relevance, provider-side unfairness has been observed to emerge in a variety of settings, ranging from content recommendations to employment websites.[116]

Because different categories of stakeholders derive different types of value from recommendation systems, issues of fairness can arise separately for each of them. In e-commerce applications, for example, users derive value from relevant recommendations for items that they might be interested in buying, while sellers derive value from their items being exposed to more potential buyers. Moreover, attempts to address unfair bias for one category of stakeholders might lead to making things worse for another category. In the case of e-commerce applications, for example, attempts to improve provider-side fairness could have negative effects on the relevance of recommendations for users. Bringing these competing interests together, comparing them and devising overarching fairness metrics remains an open challenge.[117]

Issues of fairness were not prominently mentioned by our interview participants. When fairness was referenced, it was primarily with regards to fairness concerns for users and whether recommendation systems performed better for some demographics than others. However, the extent to which recommendation systems are currently used across public service media organisations we spoke to was low enough that the risk did not generate too much concern among many staff. Sebastien Noir, European Broadcasting Union, said that 'Recommendation appears,

115 Boratto, L., Fenu, G. and Marras, M. (2021) 'Interplay between upsampling and regularization for provider fairness in recommender systems'. *User Modeling and User-Adapted Interaction*, 31(3), pp. 421–455. Available at: https://doi.org/10.1007/s11257-021-09294-8

116 Biega, A. J., Gummadi, K. P. and Weikum, G. (2018). 'Equity of Attention: Amortizing Individual Fairness in Rankings'. SIGIR '18: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 405–414. Available at: https://dl.acm.org/doi/10.1145/3209978.3210063

117 Abdollahpouri, H., Adomavicius, G., Burke, R. et al. (2020). 'Multistakeholder recommendation: Survey and research directions'. *User Modeling and User-Adapted Interaction,* pp.127–158. Available at: https://doi.org/10.1007/s11257-019-09256-1

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

72

at least for the moment more than something like [the] cherry on the cake, it's a little bit of a personalised touch on the world where everything is still pretty much broadcast content where everyone gets to receive the same content.'[118] Since, for now, recommendations represent a very small portion of the content that users access on these platforms, the risk that this poses to fairness was deemed to be very low.

However, if recommendations were to take a more prominent role in future, this would pose concerns that need to be addressed. Some of our BBC interviewees expressed a concern that some recommendations currently cater best to the interests of some demographics, while they work less well for others. Differential levels of accuracy and quality of experience across groups of users is a known issue in recommendation systems, although the way in which it manifests can be difficult to predict before the system is deployed.

In general, our respondents believed that 'majority' users, whose informational needs and preferences are closest to the average, and therefore more predictable, tend to be served best by a recommendation system – though many acknowledge this assertion has been difficult to empirically prove. If the majority of BBC users belong to a specific demographic, this could skew the system towards their interests and tastes, posing fairness issues with respect to other demographics. However, this can sometimes be reversed when other factors beyond user relevance, such as increasing the diversity of users and the diversity of content, are introduced. Therefore, the emerging patterns from recommendations are difficult to predict, but will need to be monitored on an ongoing basis. BBC interviewees reported that this issue is currently addressed by looping in more editorial oversight.

6. Social effects or externalities

One of the features of recommendation systems that has attracted most controversy in recent years is their apparent tendency to produce negative social effects. Social media networks that use recommendation systems to structure user feeds, for instance, have come under scrutiny for increasing polarisation by optimising for engagement. Other

---

118   Interview with Sébastien Noir, Head of Software, Technology and Innovation, and Dmytro Petruk, Developer, European Broadcasting
       Union (2021).

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?   73

social networks have come under fire for facilitating the spread of disinformation.

The social externality risk relates to the public service media values of **universality** (create a public sphere, in which all citizens can form their own opinions and ideas, aiming for inclusion and social cohesion) and **diversity** (support and seek to give voice to a plurality of competing views – from those with different backgrounds, histories and stories. Help build a more inclusive, less fragmented society).

Pariser introduced the concept of a 'filter bubble', which can be understood as an informational ecosystem where individuals are only or predominantly exposed to certain types of content, while they never come into contact with other types.[119] The philosopher C Thi Nguyen has offered an analysis of how filter bubbles might develop into echo chambers, where users' beliefs are reflected at them and reinforced through interaction with media that validates them, leading to potentially dangerous escalation.[120] However, some recent empirical research has cast doubt on the extent to which recommendation systems deployed on social media really give rise to filter bubbles and political polarisation in practice.[121]

In one study, it was observed that consuming news through social media increases the diversity of content consumed, with users engaging with a larger and more varied selection of news sources.[122] These studies highlight how recommendation systems can be programmed to increase the diversity of exposure to varied sources of content.[123] However, they do not control for the quality of the sources or the individual reaction to the content (e.g. does the user pay attention or merely scroll down on some of the news items?). Without this information it is difficult to know what the effects are of exposure to different types of sources. More research is needed to probe the links

119   Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. Penguin Books.

120   Nguyen, C. T. (2018). 'Why it's as hard to escape an echo chamber as it is to flee a cult'. *Aeon*. Available at:
        https://aeon.co/essays/why-its-as-hard-to-escape-an-echo-chamber-as-it-is-to-flee-a-cult

121   Arguedas, A. R., Robertson, C. T., Fletcher, R. and Nielsen R.K. (2022). 'Echo chambers, filter bubbles, and polarisation: a literature
        review.' *Reuters Institute for the Study of Journalism*. Available at:
        https://reutersinstitute.politics.ox.ac.uk/echo-chambers-filter-bubbles-and-polarisation-literature-review

122   Scharkow, M., Mangold, F., Stier, S. and Breuer, J. (2020). 'How social network sites and other online intermediaries increase exposure
        to news'. *Proceedings of the National Academy of Sciences*, 117(6), pp. 2761–2763. Available at: https://doi.org/10.1073/pnas.1918279117

123   A similar finding exists in other studies of public service media organisations – see: Hildén, J. (2021). 'The Public Service Approach
        to Recommender Systems: Filtering to Cultivate'. *Television & New Media*, 23(7). Available at:
        https://doi.org/10.1177/15274764211020106

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

74

between exposure to diverse sources and the influence this has on the evolution of political opinions.

Another known risk for recommendation systems is exposure to manipulation by external agents. Various states, for example Russia and China, have been documented to engage in what has been called 'computational propaganda'. This type of propaganda exploits some features of recommendation systems on social media to spread mis- or disinformation, with the aim of destabilising the political context of the countries targeted. State-sponsored 'content farms' have been documented to produce content that is engineered to be picked up by recommendation systems to go viral. This kind of hostile strategy is made possible by the vulnerability of the recommendation system, especially open ones, because the system is programmed to optimise for engagement.

The risk that the use of recommendation systems could increase polarisation and create filter bubbles was regarded as very low by our interviewees. Unlike social media that recommend content generated by users or other organisations, the BBC and other public service media that we spoke to operate closed content platforms. This means that all the content recommended on their platforms has already passed multiple editorial checks, including for balanced and truthful reporting.

The relatively minor role that recommendation systems play on the platform currently also means that they do not pose a risk of creating filter bubbles. Therefore, this was not recognised as a pressing concern.

However, many raised concerns that recommendation systems could undermine the principle of diversity by serving audiences homogenous content. Historically, programme schedulers have had mechanisms to expose audiences to content they might not choose of their own accord – for example by 'hammocking' programmes of high public value between more popular items on the schedule and relying on audiences not to switch channels. Interviewees also mentioned the importance of serendipity and surprise as part of the public service remit. This could be lost if audiences are only offered content based on their previous preferences. These concerns motivate ongoing research into new methods for producing more accurate and diversified

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

75

recommendations.[124]

Conclusion

The categories of risk related to the use of recommendation systems, identified in the literature, can be applied to their use in the context of public service media. However, the way in which these risks manifest and the emphasis that organisations put on them can be quite different to a commercial context.

We found that public service media have, to a greater or lesser extent, mitigated their exposure to these risks through a number of factors such as the high quality of the content being recommended; the limited deployment of the systems; the substantial level of human curation; a move towards greater integration of technical and editorial teams; ethical principles; associated practice checklists and system documentation. It is not enough for public service media organisations to believe that having a public service mission will ensure that recommendation systems serve the public. If public service media are to use recommendation systems responsibly, they must interrogate and mitigate the potential risks.

We find these risks can also be seen in relation to the six core public service values of **universality**, **independence**, **excellence**, **diversity**, **accountability** and **innovation**.

> We believe it is useful for public service media to consider both the known risks, as understood within the wider research field, as well as the risks in relation to public service values. By approaching the potential challenges of recommendation systems through this dual lens, public service media organisations should be able to develop and deploy systems in line with their public service remit.

124 Paudel, B., Christoffel, F., Newell, C. and Bernstein, A. (2017). 'Updatable, Accurate, Diverse, and Scalable Recommendations for Interactive Applications'. *ACM Transactions on Interactive Intelligent Systems*, 7(1), pp.1–34. Available at: https://doi.org/10.1145/2955101

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

76

An additional consideration, broader than any specific risk category, is that of audience trust in public service media. Trust doesn't fall under any specific category because it is associated with the relationship between public service media and their audience more broadly. But failure to address the risks identified by the categories can negatively affect trust. All public service media organisations place trust as central to their mission. In the context of a fragmented digital media environment, their trustworthiness has taken on increased importance and is now a unique quality that distinguishes them from other media and which is pivotal to the argument in favour of sustaining public service media. Many public service media organisations are beginning to recognise and address the potential risks of recommendation systems and it is vital that this continues in order to retain audience trust.

Additional challenges for public service media

As well as the ethical risks described above, public service media face practical challenges in implementing recommendation systems that stem from their mission, the make-up of their teams and their organisational infrastructure.

Quantifying values

Recommendation systems filter content according to criteria laid down by the system developers. Public service media organisations that want to filter content in ways that prioritise public service values first need to translate these values into information that is legible to an algorithmic system. In other words, the values must be quantified as data.

However, as we noted above, public service values are fluid, can change over time and depend on context. And as well as the stated mission of public service media, laid down in charters, governance and guidelines, there are a set of cultural norms and individual gut instincts that determine day-to-day decision making and prioritisation in practice. Over time, public service media have developed a number of ways to measure public value, through systems such as the public value test assessment and with metrics such as audience reach, value for money and surveys of public sentiment (see section above). However, these only account for public value at a macro level. Recommendation systems that are

filtering individual items of content require metrics that quantify values at a micro level.

Swedish Radio is a pioneer in attempting to do this work of translation. Olle Zachrison of Swedish Radio summarised it as: 'we have central tenets to our public service mission stuff that we have been talking about for decades and also stuff that is in the kind of gut of the news editors. But in a way, we had to get them out there in an open way and into a system also, that we in a way could convert those kinds of editorial values that have been sitting in these kind of really wise news assessments for years, but to get them out there into a system that we also convert them into data.'[125]

Working across different teams and different disciplines

The development and deployment of recommendation systems for public service media requires expertise in both technical development and content creation and curation. This proves challenging in a number of ways.

Firstly, technology talent is hard to come by, especially when public service media cannot offer anything near the salaries available at commercial rivals.[126] Secondly, editorial teams often do not trust or value the role of technologists, especially when the two do not work closely with each other.[127] In some organisations, the introduction of recommendation systems stalls because it is perceived as a direct threat to editorial jobs and an attempt to replace journalists with algorithms.[128]

Success requires bridging this gap and coordinating between teams of experts in technical development, such as developers and data scientists, and experts in content creation and curation, the journalists and editors.[129]

125 Interview with Olle Zachrison, Deputy News Commissioner & Head of Digital News Strategy, Swedish Radio (2021).

126 Interview with Dietmar Jannach, Professor, University of Klagenfurt (2021).

127 Interview with Nic Newman, Senior Research Associate, Reuters Institute for the Study of Journalism (2021).

128 Interview with Sébastien Noir, Head of Software, Technology and Innovation, and Dmytro Petruk, Developer, European Broadcasting Union (2021).

129 Boididou, C., Sheng, D., Moss, M. and Piscopo, A. (2021), 'Building Public Service Recommenders: Logbook of a Journey'. *RecSys '21: Proceedings of the 15th ACM Conference on Recommender Systems*, pp. 538–540. Available at: https://doi.org/10.1145/3460231.3474614

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain… and recommend?

78

As Sørensen and Hutchinson note: 'Data analysts and computer programmers (developers) now perform tasks that are key determinants for exposure to public service media content. Success is no longer only about making and scheduling programmes. This knowledge is difficult to communicate to journalists and editors, who typically don't engage in these development projects […] Deep understanding of how a system recommends content is shared among a small group of experts'.[130]

Some, such as Swedish Radio and BBC News Labs, have tried to tackle this issue by explicitly having two project leads, one with an editorial background and one with a technical background, to emphasise the importance of working together and symbolically indicate that this was a joint process.[131] Swedish Radio's Olle Zachrison noted that:

> 'We had a joint process from day one. And we also deliberately had kind of two project managers, one, clearly from the editorial side, like a very experienced local news editor. And the other guy was the product owner for our personalization team. So they were the symbols internally of this project […] that was so important for the, for the whole company to kind of team up behind this and also for the journalists and the product people to do it together.'

If this coordination fails, this can 'weaken the organisation strategically and, on a practical level, create problems caused by failing to include or correctly mark the metadata that is essential for findability'.

Bayerische Rundfunk has established a unique interdisciplinary team. The AI and Automation Lab has a remit to not only create products, but also produce data-driven reporting and coverage of the impacts of artificial intelligence on society. Building from the existing data journalism unit, the Lab fully integrates the editorial and technical teams under the leadership of Director Uli Köppen. Although she recognises the challenges of bringing together people from different backgrounds, she believes the effort has paid off:

130   Sørensen, J.K. and Hutchinson, J. (2018). 'Algorithms and Public Service Media'. *Public Service Media in the Networked Society: RIPE@2017*, pp.91–106. Available at:
      http://www.nordicom.gu.se/sites/default/files/publikationer-hela-pdf/public_service_media_in_the_networked_society_ripe_2017.pdf
131   Interview with Olle Zachrison, Deputy News Commissioner & Head of Digital News Strategy, Swedish Radio (2021); BBC News Labs. 'About'. Available at: https://bbcnewslabs.co.uk/about

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?   79

'This technology is so new, and it's so hard to persuade the experts to work in journalism. We had the data team up and running, these are journalists that are already in the mindset at this intersection of tech and journalism. And I had the hope that they are able to help people from other industries to dive into journalism, and it's easier to have this kind of conversation with people who already did this cultural step in this hybrid world.

'It was astonishing how those journalists helped the new people to onboard and understand what kind of product we are. And we are also reinventing our role as journalists in the product world. And this really worked out so I would say it's worth the effort.'

Metadata, infrastructure and legacy systems

In order to filter content, recommendation systems require clear information about what that content is. For example, if a system is designed to show people who enjoyed soap operas other series that they might enjoy, individual items of content must be labelled as being soap operas in a machine-readable format. This kind of labelling is called metadata.

However, public service media have developed their programming around the needs of individual channels and stations organised according to particular audiences and tastes (e.g. BBC Radio 1 is aimed at a younger audience around music, BBC Radio 4 at an older audience around speech content) or by a particular region (e.g. in Germany Bayerische Rundfunk serves Bavaria, WDR serves West Germany but both are members of the federal broadcaster ARD). Each of these channels will have evolved their own protocols and systems and may label content differently – or not at all. This means the metadata to draw on for the deployment of recommendation systems is often sparse and low quality, and the metadata infrastructure is often disjointed and unsystematic.

We heard from many interviewees across public service media organisations that access to high-quality metadata was one of the most significant barriers to implementing recommendation systems. This was particularly an issue when they wanted to go beyond the most simplistic approaches and experiment with assigning public service value to pieces of content or measuring the diversity of recommended content.

Stated goals and
potential risks of using
recommendation systems
in public service media

Inform, educate, entertain... and recommend?

80

Recommendation system projects often required months of setting up systems for data collection, then assessing and cleaning that data, before the primary work of building a recommendation system could begin. To achieve this requires a significant strategic and financial commitment on the part of the organisation, as well as buy-in from the editorial teams involved in labelling.

Evaluation of recommendation systems

We've explored the possible benefits and harms of recommendation systems, and how those benefits and harms might manifest in a public service media context. To try to understand whether and when those benefits and harms occur, developers of recommendation systems need to evaluate their systems. Conversely, looking at how developers and organisations evaluate their recommendation systems can tell us what benefits and harms, and to whom, they prioritise and optimise for in their work.[132]

In this chapter, we look at:

- how recommendation systems can be evaluated
- how public service media organisations evaluate their own recommendation systems
- how evaluation might be done differently in future.

How recommendation systems are evaluated

In this section, we lay out a framework for understanding the evaluation of recommendation systems as a three-stage process of:

1. Setting objectives.
2. Identifying metrics.
3. Selecting methods to measure those metrics.

---

132   Evaluation of recommendation systems in not limited to the developers and deployers of those systems. Other stakeholders such as users, government, regulators, journalists and civil society organisations may all have their own goals for what they think a particular recommendation system should be optimising for. Here however, we focus on evaluation as seen by the developer and deployer of the system, as this is where there is the tightest feedback loop between evaluation and changes to the system and the developers and deployers generally have privileged access to information about the system and a unique ability to run tests and studies on the system. For more on how regulators (and others) can evaluate social media companies in an online-safety context, see: Ada Lovelace Institute. (2021). *Technical methods for regulatory inspection of algorithmic systems*. Available at: https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/

This framework is informed by three aspects of evaluation (objectives, metrics and methods) as identified by Francesco Ricci, Professor of Computer Science at the Free University of Bozen-Bolzano.

Objectives

Evaluation is a process of determining how well a particular system achieves a particular set of goals or objectives. To evaluate a system, you need to know what goals you are evaluating against.[133]

However, this is not a straightforward exercise. There is no singular goal for a recommendation system and different stakeholders will have different goals for the system. For example, on a privately-owned social media platform:

- the engineering team's goal might be to create a recommendation system that serves 'relevant' content to users
- the CEO's goal might be to maximise profit while minimising personal reputational risk
- the audience's goal may be to discover new and unexpected content (or just avoid boredom).

If a developer wants to take into account the goals of all the stakeholders in their evaluation, they will need to decide how to prioritise or weigh these different goals.[134]

Balancing goals is ultimately a 'political' or 'moral' question, not a technical one, and there will never be a universal answer about how to weigh these different factors, or even who the relevant stakeholders whose goals should be weighted are.

> Any process of evaluation ultimately needs a process to determine the relevant stakeholders for a recommendation system and how their priorities should be weighted.

133 Interview with Francesco Ricci, Professor of Computer Science, Free University of Bozen-Bolzano (2021).
134 Interview with Francesco Ricci.

This is made more difficult because people are often confused or uncertain about their goals, or have multiple competing goals, and so the process of evaluation will need to help people clarify their goals and their own internal weightings between those goals.

Metrics

Furthermore, goals are often quite general and whether they have been met cannot be directly observed.[135] Therefore, once a goal has been decided, such as 'relevance to the user', the goal needs to be operationalised into a set of specific metrics to judge the recommendation system against.[136] These metrics can be quantitative, such as the number of users who click on an item, or qualitative, such as written feedback from users about how they feel about a set of recommendations.

Whatever the metrics used, the choice of metrics is always a choice of a particular interpretation of the goal. The metric will always be a proxy for the goal, and determining a proxy is a political act that grants power to the evaluator to decide what metrics reflect their view of the problem to be solved and the goals to be achieved.[137]

The people who define these metrics for the recommendation system are often the engineering or product teams. However, these teams are not always the same people who set the goals of an organisation. Furthermore, they may not directly interact with other stakeholders who have a role in setting the goals of the organisation or the goal of deploying the recommendation system.

Therefore, through misunderstanding, lack of knowledge or lack of engagement with others' views, the engineering and product teams' interpretation of the goal will likely never quite match the intention of the goal as envisioned by others.

---

135 Interview with Francesco Ricci, Professor of Computer Science, Free University of Bozen-Bolzano (2021).

136 Operationalising is a process of defining how a vague concept, which cannot be directly measured, can nevertheless be estimated by empirical measurement. This process inherently involves replacing one concept, such as 'relevance', with a proxy for that concept, such as 'whether or not a user clicks on an item' and thus will always involve some degree of error.

137 Beer, D. (2016). *Metric Power.* London: Palgrave Macmillan. Available at: https://doi.org/10.1057/978-1-137-55649-3

Metrics will also always be a simplified vision of reality, summarising individual interactions with the recommendation system into a smaller set of numbers, scores or lines of feedback.[138] This does not mean metrics cannot be useful indicators of real performance; this very simplicity is what makes them useful in understanding the performance of the system. However, those creating the metrics need to be careful not to confuse the constructed metric with the reality underlying the interactions of people with the recommendation system. The metric is a measure of the interaction, not the interaction itself.

Methods

Evaluating is then the process of measuring these metrics for a particular recommendation system in a particular context, which requires gathering data about the performance of the recommendation system. Recommendation systems are evaluated in three main ways:[139]

1. **Offline evaluations** test recommendation systems without real users interacting with the system, for example by measuring recommendation system performance on historical user interaction data or in a synthetic environment with simulated users.

2. **User studies** test recommendation systems against a small set of users in a controlled environment with the users being asked to interact with the system and then typically provide explicit feedback about their experience afterwards.

3. **Online evaluations** test recommendation systems deployed in a live environment, where the performance of the recommendation system is measured against interactions with real users.

These methods of evaluation are not mutually exclusive and a recommendation system might be tested with each method sequentially, as it moves from design to development to deployment.

138 Raji, I. D., Bender, E. M., Paullada, A. et al. (2021). 'AI and the Everything in the Whole Wide World Benchmark', p2. arXiv. Available at: https://doi.org/10.48550/arXiv.2111.15366

139 Gunawardana, A. and Shani, G. (2015). 'Evaluating Recommender Systems'. *Recommender Systems Handbook*, pp 257–297. Available at: https://doi.org/10.1007/978-0-387-85820-3_8

Offline evaluation has been a historically popular way to evaluate recommendation systems. It is comparatively easy to do, due to the lack of interaction with real users or a live platform. In principle, they are reproducible by other evaluators, and allow standardised comparison of the results of different recommendation system.[140]

However, there is increasing concern that offline evaluation results based on historical interaction data do not translate well into real-world recommendation system performance. This is because the training data is based on a world without the new recommendation system in it, and evaluations therefore cannot account for how that system might itself shift wider aspects of the service like user preferences.[141] This limits their usefulness in evaluating which recommendation system would actually be the best performing in the dynamic live environments most stakeholders are interested in, such as a video-sharing website with an ever-growing set of videos and ever-changing set of viewers and content creators.

Academics we spoke to in the field of recommendation systems identified user studies in labs and simulations as the state of the art in academic recommendation system evaluation. Whereas in industry, common practice is to use online evaluation via A/B testing to optimise key performance indicators.[142]

How do public service media evaluate their recommendation systems?

In this section, we use the framework of objectives, metrics and methods to examine how public service media organisations evaluate their recommendation systems in practice.

140 Jannach, D. and Jugovac, M. (2019), 'Measuring the Business Value of Recommender Systems'. *ACM Transactions on Management Information Systems*, 10(4), pp 1–23. Available at: https://doi.org/10.1145/3370082

141 Rohde, D., Bonner, S., Dunlop, T., et al. (2018). 'RecoGym: A Reinforcement Learning Environment for the problem of Product Recommendation in Online Advertising'. *arXiv*. Available at: https://doi.org/10.48550/arXiv.1808.00720; Beel, J. and Langer, S. (2015). 'A Comparison of Offline Evaluations, Online Evaluations, and User Studies in the Context of Research-Paper Recommender Systems'. *Proceedings of the 19th International Conference on Theory and Practice of Digital Libraries (TPDL)*, pp.153-168. Available at: doi: 10.1007/978-3-319-24592-8_12; Jannach, D., Pu, P., Ricci, F. and Zanker, M. (2021). 'Recommender Systems: Past, Present, Future'. *AI Magazine*, 42 (3). Available at: https://doi.org/10.1609/aimag.v42i3.18139

142 Interview with Dietmar Jannach, Professor, University of Klagenfurt (2021).

Objectives

As we discussed in the previous chapter, recommendation systems are ultimately developed and deployed to serve the goals of the organisation using them; in this case, public service media organisations. In practice, however, the objectives that recommendation systems are evaluated against are often multiple levels of operationalisation and contextualisation down from the overarching public service values of the organisation.

For example, as discussed previously, the BBC Charter agreement sets out the mission and public purposes of the organisation for the following decade. These are derived from the public service values, but are also shaped by political pressures as the Charter is negotiated with the British Government of the time.

The BBC then publishes an annual plan setting out the organisation's strategic priorities for that year, drawing explicitly on the Charter's mission and purposes. These annual plans are equally shaped by political pressures, regulatory constraints and challenges from commercial providers. The plan also sets out how each product and service will contribute towards meeting those strategic priorities and purposes, setting the goals for each of the product teams.

For example, the goals of BBC Sounds as a product team in 2021 were to:

1. Increase the audience size of BBC Sounds' digital products.
2. Increase the demographic breadth of consumption across BBC Sounds' products, especially among the young.
3. Convert 'lighter users' into regular users.
4. Enable users to more easily discover content from the more than 50 hours of new audio produced by the BBC on an hourly basis.[143]

These objectives map onto the goals for using recommendation systems we discussed in the previous chapter. Specifically, the first three relate

---

143 According to David Jones (Executive Product Manager, BBC Sounds, interviewed in 2021), his top-line KPI is to reach 900,000 members of the British population who are under 35 by March 2022. These numbers are determined centrally by BBC senior managers based on the BBC's Service Licence for BBC Online and Red Button. See: BBC Trust. (2016). *BBC Online and Red Button Service Licence*. Available at:
http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/regulatory_framework/service_licences/online/2016/online_red_button_may16.pdf

to capturing audience attention and the fourth relates to reducing information overload and improving discoverability for audiences.

These product goals then inform the objectives of the engineering and product teams in the development and deployment of a recommendation system, as a feature within the wider product.

> At each stage, as the higher level objectives are interpreted and contextualised lower down, they may not always align with each other.

The objectives for the development and deployment of recommendation systems in public service media seem most clear for entertainment products, e.g. audio-on-demand and video-on-demand. Here, the goal of the system is clearly articulated as a combination of audience engagement with reaching underserved demographics and serving more diverse content. These are often explicitly linked by the development teams to achieving the public service values of diversity and a personalised version of universality, which they see as serving the needs of each and every group in society

In these cases, public service media organisations seem better at articulating goals for recommendation systems when they are using recommendation systems for a similar purpose as private-sector commercial media organisations. This seems, in part, because there is greater existing knowledge of how to operationalise those objectives, and the developers can draw on their own private sector experience and existing industry practice, open-source libraries and similar resources.

However, when setting objectives that focus more focus on public service value, public service media organisations often seem less clear about the goals of the recommendation system within the wider product.

This seems partly because in the domain of news, for example, the use of recommendation systems by public service media is more experimental and at an earlier stage of maturity. Here, the motivations often come further apart from commercial providers, with the implicit motivation of public service media developers seemingly to augment existing editorial capabilities with a recommendation system, rather than drive engagement with the news content. This means public service media

developers have less existing practices and resources to draw upon for translating product goals and articulating recommendation system objectives in those domains.

In general, it seems that some public service values are easier to operationalise in the context of recommendation systems than others, such as diversity and universality. These values get privileged over others, such as accountability, in the development of recommendation systems, as they are the easiest to translate through from the overarching set of organisational values down to the product and feature objectives.

Metrics

Public service media organisations have struggled to operationalise their complex public service values into specific metrics. There seem to be three broad responses to this:

1. Fall back on established engagement metrics, e.g. click-through rate and watch time, often with additional quantitative measures of the diversity of audience content consumption.

2. The above approach combined with attempts to create crude numerical measures (e.g. a score from 1 to 5) of 'public service value' for pieces of content, often reducing complex values to a single number subjectively judged by journalists, then measuring the consumption of content with a 'high' public service value score.

3. Try to indirectly optimise for public service value by making their metrics the satisfaction of editorial stakeholders, whose preferences are seen as the best 'ground truth' proxy for public service value. Then optimise for lists of recommendations which are seen to have high public service value by editorial stakeholders.

Karin van Es found that, as of 2017, the European Broadcasting Union and the Dutch public service media organisation NPO evaluated pilot algorithms using the same metrics found in commercial systems i.e.

stream starts and average-minute ratings.[144] As van Es notes, these metrics are a proxy for audience retention and even if serving diverse content was an explicit goal in designing the system, the chosen metrics reflect – and will ultimately lead to – a focus on engagement over diversity. Therefore, despite different stated goals, the public service media use of recommendation systems ends up optimising for similar outcomes as private providers.

By now, most public service media organisations using recommendation systems also have explicit metrics for diversity, although there is no single shared definition of diversity across the different organisations, nor is there one single metric used to measure the concept.

However, most quantitative metrics for diversity in the evaluation of public service media recommendation systems focus on diversity in terms of audience exposure to unique pieces of content or to categories of content, rather than on the representation of demographic groups and viewpoints across the content audiences are exposed to.[145]

Some aspects of diversity, as Hildén observes, are easier to define and 'to incorporate into a recommender system than others. For example, genres and themes are easy to determine at least on a general level, but questions of demographic representation and the diversity of ideas and viewpoints are far more difficult as they require quite detailed content tags in order to work. Tagging content and attributing these tags to users might also be politically sensitive especially within the context of news recommenders'.[146]

Commonly used metrics for diversity include intra-list diversity, i.e. the average difference between each pair of items in a list of recommendations and inter-list diversity, i.e. the ratio of items recommended to total items recommended across all the lists of recommendations.

144 van Es, K. F. (2017). 'An Impending Crisis of Imagination : Data-Driven Personalization in Public Service Broadcasters'. *Media@LSE*. Available at: https://dspace.library.uu.nl/handle/1874/358206

145 This was generally attributed by interviewees to a combination of a lack of metadata to measure the representativeness within content and assumption that issues of representation within content were better dealt with at the point at which content is commissioned, so that the recommendation systems have diverse and representative content over which to recommend.

146 Hildén, J. (2021). 'The Public Service Approach to Recommender Systems: Filtering to Cultivate'. *Television & New Media*, 23(7). Available at: https://doi.org/10.1177/15274764211020106

Some public service media organisations are experimenting with more complex measures of exposure diversity. For example, Koen Muylaert at Belgian VRT explained how they measure an 'affinity score' for each user for each category of content, e.g. your affinity with documentaries or with comedy shows, which increases as you watch more pieces of content in that category.[147] VRT then measure the diversity of content that each user consumes by looking at the difference between a user's affinity scores for different categories.[148] VRT see this method of measuring diversity as valuable because they can explain it to others and measure it across users over time, to track how new iterations of their recommendation system increase users' exposure to diverse content.

To improve on this, some public service media organisations have tried to implement 'public service value' as an explicit metric in evaluating their recommendation systems. NPO, for example, ask a panel of 1,500 experts and ordinary citizens to assess the public value of each piece of content, including the diversity of actors and viewpoints represented in the content, and then ask those panellists to assign a single 'public value' from 1 to 100 to all pieces of content on their on-demand platform. They then calculate an average 'public value' score for the consumption history of each user. According to Sara van der Land, Digital Innovation Advisor at NPO, their target is to make sure that the average 'public value' score of every user rises over time.[149]

At the moment, they are only specifically focusing on optimising for that metric within a specific 'public value' recommendations section within their wider on-demand platform, which is a mixture of recommendations based on user engagement and the 'public value' of the content. However, through experiments, they found there was a trade-off between optimising for 'public value' and viewership, as noted by Arno van Rijswijk, Head of Data & Personalization at NPO:

147 Interview with Koen Muylaert, Project Lead, VRT data platform and data science initiative, Vlaamse Radioen Televisieomroeporganisatie (VRT) (2021).

148 By measuring the entropy of the distribution of affinity scores across categories, and trying to improve diversity by increasing that entropy.

149 Interview with Arno van Rijswijk, Head of Data & Personalization, and Sarah van der Land, Digital Innovation Advisor, Nederlandse Publieke Omroep (2021).

> 'When we're focusing too much on the public value, we see that the percentage of people that are watching the actual content from the recommender is way lower than when you're using only the collaborative filtering algorithm […] So when you are focusing more on the relevance then people are willing to watch it. And when you're adding too much weight on the public values, people are not willing to watch it anymore.'

This resulted in them choosing to have a 'low ratio' of public value content to engaging content, making explicit the choice that public service media organisations often do and have to make between audience retention and other public service values like diversity, at least over the short-term these metrics measure.

Others, when faced with the inadequacy of conventional engagement and diversity metrics, have tried to indirectly optimise for public service value by making their metrics the satisfaction of editorial stakeholders, whose preferences are seen as the best 'ground truth' proxy for public service value.

In the early stages of developing an article-to-article news recommendation system in 2018,[150] the BBC Datalab initially used a number of quantitative metrics for its offline evaluation.[151] They evaluated these using offline metrics, with proxies for engagement, diversity and relevance to audiences, including:

- hit rate, i.e. whether the list of recommended articles includes an article a user did in fact view within 30 minutes of viewing the original article
- normalised discounted cumulative gain, i.e. how relevant the recommended articles were assumed to be to the user, with a

---

150   The Datalab team was experimenting with and evaluating a number of approaches using a combination of content and user interaction data, such as neural network approaches that combine both content and user data as well as collaborative filtering models based only on user interactions.

151   Panteli, M., Piscopo, A., Harland, A., Tutcher, J. and Moss, F. M. (2019). 'Recommendation systems for news articles at the BBC', p. 4. *CEUR Workshop Proceedings*. Available at: http://ceur-ws.org/Vol-2554/paper_07.pdf

higher weighting for the relevance of articles higher up in the list of
recommendations

- intra-list diversity, i.e. the average difference between every pair of
articles in a list of recommendations

- inter-list diversity, i.e. the ratio of unique articles recommended to total
articles recommended across all the lists of recommendations

- popularity-based surprisal, i.e. how novel the articles recommended
were

- recency, i.e. how old the articles recommended were when shown to
the user.

However, they found that performance on these metrics didn't match
the editorial teams' priorities. When they tried to instead operationalise
into metrics what public service value meant to the editors, existing
quantitative metrics were unable to capture editorial preferences and
creating new ones was not straightforward. As Alessandro Piscopo, Lead
Data Scientist, BBC Datalab notes:[152]

> 'We did notice that in some cases, one of the recommender prototypes
> was going higher in some metrics and went to editorial and [they
> would] say well we just didn't like it [...] Sometimes it was just comments
> from editorial world, we want to see more depth. We want to see more
> breadth. Then you have to interpret what that means.'

This difficulty in finding appropriate metrics led to the Datalab team
changing their primary method of evaluation, from offline evaluation
to user studies with BBC editorial staff, which they called 'subjective
evaluation'.[153]

In this approach, they asked editorial staff to score each list of articles
generated by the recommendation systems as either: unacceptable,
inappropriate, satisfactory or appropriate. The editors were then
prompted to describe what properties they considered in choosing
how appropriate the recommendations were. The development team
would then iterate the recommendation system based on the scoring
and written feedback along with discussion with editorial about the
recommendation.

152 Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

153 Piscopo, A. (2021). 'Building public service recommenders: Logbook of a journey' [presentation recording]. *The Academic Fringe
Festival*. Available at: https://www.youtube.com/watch?v=Q2EYAxX5Pnk

Early in the process, the Datalab team agreed with editorial what percentage of each grade they were aiming for, and so what would be a benchmark for success in creating a good recommendation system. In this case, the editorial team decided that they wanted:[154]

1.   No unacceptable recommendations, on the basis that any unacceptable recommendations would be detrimental to the reputation of the BBC.

2.   Maximum 10% inappropriate recommendations.

This change of metrics meant that the evaluation of the recommendation system, and the iteration of the system as a result, was optimising for the preferences of the editorial team, over imperfect measures of audience engagement, relevance and diversity. The editors are seen as the most reliable 'source of truth' for public service value, in lieu of better quantitative metrics.

Methods

Public service media often rely on internal user studies with their own staff as an evaluation method during the pre-deployment stage of recommendation system development. For example, Greg Detre, ex-Chief Data Scientist at Channel 4, said that when developing a recommendation system for All 4 in 2016, they would ask staff to subjectively compare the output of two recommendation systems side by side, based on the staff's understanding of Channel 4's values:

'So we're making our recommendations algorithms fight, "Robot Wars" style, pick the one that you think […] understood this view of the best, good recommendations are relevant and interesting to the viewer. Great recommendations go beyond the obvious. Let's throw in something a little unexpected, or showcase the Born Risky programming that we're most proud of, [clicking the] prefer button next to the […]one you like best […] Born Risky, which was one of the kind of Channel Four cultural values for like, basically being a bit cheeky. Going beyond the mainstream, taking a chance. It was one of, I think, a handful of company values.'[155]

---

154   Piscopo, A. (2021); Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

155   Interview with Greg Detre, ex-Chief Data Scientist, Channel 4 (2021).

Similarly, when developing a recommendation system for BBC Sounds, the BBC Datalab decided to use a process of qualitative evaluation. BBC Sounds uses a factorisation machine approach, which is a mixture of content matching and collaborative filtering. This uses your listening history, metadata about the content and other users' listening history to make recommendations in two ways:

1. It recommends items that have similar metadata to items you have already listened to.

2. It recommends items that have been listened to by people with otherwise similar listening histories.

When evaluating this approach, BBC compared the new factorisation machine recommendation system head-to-head with the existing external provider's recommendations.

They recruited 30 BBC staff members under the age of 35 to be test users.[156] They then showed these test users two sets of nine recommendations side by side. One set was provided by the current external provider's recommendation system, and the other set was provided by the team's internal factorisation machine recommendation system. The users were not told which system had produced which set of recommendations, and had to choose whether they preferred 'A' or 'B', or 'both' or 'neither', and then explain their decision why in words.

Over 60% of test users preferred the recommendation sets provided by the internal factorisation machine.[157] This convinced the stakeholders that the system should move into production and A/B testing, and helped editorial teams get hands-on experience evaluating automated curations, increasing their confidence in the recommendation system.

Similarly, when later deploying the recommendation system to create personalised sorting system for feature items, the Datalab team held a number of digital meetings with editorial staff, showing them the personalised and non-personalised featured items side-by-side. The

156 Al-Chueyr Martins, T. (2021). 'From an idea to production: the journey of a recommendation engine' [presentation recording]. *MLOps London*. Available at: https://www.youtube.com/watch?v=dFXKJZNVgw4

157 Al-Chueyr Martins, T. (2021).

Datalab then got feedback from the editors on which they preferred.[158] This approach allowed them to more directly capture internal staff preferences and manually step towards meeting those preferences. However, the team acknowledged its limitations upfront, particularly in terms of scale.[159] Editorial teams and other internal staff only have so much capacity to judge recommendations, and thus would struggle to assess every edge case or judge recommendations, if every recommendation changed depending on the demographics of the audience member viewing it.

Once the recommendation systems are deployed to a live environment, i.e. accessible by audiences on their website or app, public service media all have some form of online evaluation in place, most commonly in the form of A/B testing in which viewers are given two different recommendations to choose from.

Channel 4 used online evaluation in the form of A/B testing to evaluate the recommendation system used by their video-on-demand service, All 4. Greg Detre noted that:

> 'We did A/B test it eventually. And it didn't show a significant effect. That said [Channel 4] had an already somewhat good system in place. That was okay. And we were very constrained in terms of the technical solutions that we were allowed, there were only a very, very limited number of algorithms that we were able to implement, given the constraints that have already been agreed when I got there. And so as a result, the solution we came up with was, you know, efficient in terms of it was fast to compute in real time, and easy to sort of deploy, but it wasn't that great... I think perhaps it didn't create that much value.' [160]

BBC Datalab also used A/B testing in combination with continued user studies and behavioural testing. By April/May 2020, editorial had given sign-off and the recommendation system was deemed ready for initial deployment. [161]

158   Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

159   Interview with Alessandro Piscopo.

160   Interview with Greg Detre, ex-Chief Data Scientist, Channel 4 (2021).

161   Piscopo, A. (2021). 'Building public service recommenders: Logbook of a journey' [presentation recording]. *The Academic Fringe Festival*. Available at: https://www.youtube.com/watch?v=Q2EYAxX5Pnk

During deployment, the team took a 'failsafe approach' with weekly monitoring of the live version of the recommendation system by editorial staff. This included further subjective evaluation described above and behavioural tests. In these behavioural tests, developers use a list of pairs of inputs and desired outputs, comparing the output of the recommendation system with the desired output for each given input.[162]

After deployment, there was still a need to understand the effect and success of the recommendation systems. This took the form of A/B testing the live system. This included measuring the click-through rate on the recommended articles. However, members of the development team noted it was only a rough proxy for user satisfaction and were working to go beyond click-through rate.

Ultimately at the post-deployment stage, the success of the recommendation system is determined by the product teams, with input by development teams in the identification of appropriate metrics. It is editorial considerations that are central to product teams decide which metrics they think they are best suited to evaluate for.[163]

Once the system reaches the stage of online evaluation, these methods can only tell public service media whether the recommendation system was worthwhile after it is has already been built and considering the time and resources invested in building it. Therefore the evaluation becomes about whether to continue to use and maintain the system given the operating costs versus the costs involved in removing or replacing it. This can mean even systems that only provide limited value to the audience or to the public service media organisation will remain in use in this phase of evaluation.

How could evaluations be done differently?

In this section, we explore how the objectives, metrics and methods for evaluating recommendation systems could be done differently by public service media organisations.

162 See: BBC. (2022). *RecList*. GitHub. Available at: https://github.com/bbc/datalab-reclist; Tagliabue, J. (2022). 'NDCG Is Not All You Need'. *Towards Data Science*. Available at: https://towardsdatascience.com/ndcg-is-not-all-you-need-24eb6d2f1227

163 Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

Objectives

Some public service media organisations could benefit from more explicitly drawing a connection from their public service values to the organisational and product goals and finally to the recommendation system itself, showing how each level links to the next. This can help prevent value drift as goals go through several levels of interpretation and operationalisation, and help contextualise the role of the recommendation system in achieving public value within the wider process of content delivery.

More explicitly connecting these objectives can help organisations to recognise that, while a product as a whole should achieve public service objectives, a recommendation system doesn't need to achieve every objective in isolation. While a recommendation system's objectives should not be in conflict with the higher level objectives, they may only need to achieve some of those goals (e.g. its primary purpose might be to attract and engage younger audiences and thus promote diversity and universality). Therefore, its contribution to the product and organisational objectives should be seen in the context of the overall audience experience and the totality of the content an individual user interacts with. Evaluating against the recommendation system's feature-level objectives alone is not enough to know whether a recommendation system is also consistent with product and organisational objectives.

Audience involvement in goal-setting

Another area worthy of further exploration is providing greater audience input and control over the objectives and therefore the initial system design choices. This could involve eliciting individual preferences from a panel of audience members and then working with staff to collaboratively trade-off and explicitly set different weighting for different objectives of the system. This should take place as part of a broader co-design approach at the product level. This is because the evaluation process for a recommendation system should include the option to say a recommendation system is not the most appropriate tool for achieving the higher-level objectives of the product and providing the outcomes the staff and the audiences want from the product, rather than constraining audiences to just choose between different versions of a recommendation system.

Making safeguards an explicit objective in system evaluation

A final area worthy of exploration is building in system safeguards like accountability, transparency and interpretability as explicit objectives in the development of the system, rather than just as additional governance considerations. Some interviewees suggested making considerations such as interpretability a specific objective in evaluating recommendation systems. By explicitly weighing those considerations against other objectives and attempting to measure the degree of interpretability or transparency, it would ensure greater salience of those safeguards in the selection of systems.[164]

Metrics

More nuanced metrics for public service value

If public service media organisations want to move beyond optimising for a mix of engagement and exposure diversity in their recommendation systems, then they will need to develop better metrics to measure public service value. As we've seen above, some are already moving in this direction with varying degrees of success, but more experimentation and learning will be required.

> When creating metrics for public service value, it will be important to disambiguate between different meanings of 'public service value'. A public service media organisation cannot expect to have one quantitative measure of 'public service value', which conflates a number of priorities that can be in tension with one another.

One approach would be to explicitly break each public service value down into separate metrics for universality, independence, excellence, diversity, accountability and innovation, and most likely sub-values within those. This could help public service media developers to clearly

---

164   Interview with Greg Detre, ex-Chief Data Scientist, Channel 4 (2021).

articulate the components of each value and make it explicit how they are weighted against each other. However, quantifying concepts like accountability and independence can be challenging to do, and this approach may struggle to work in practice. More experimentation is needed.

The most promising approach may be to adopt more subjective evaluations of recommendation systems. This approach recognises that 'public service value' is going to be inherently subjective and uses metrics which reflect that. Qualitative metrics based on feedback from individuals interacting with the recommendation system can let developers balance the tensions between different aspects of public service value. This places less of a burden on developers to weight those values themselves, which they might be poorly suited to, and can accommodate different conceptions of public service value from different stakeholders.

However, subjective evaluations do have their limits. They are only able to evaluate a tiny subset of the overall recommendations, and will only capture the subjective evaluation of features appearing in that subset. These evaluations may miss features that were not present in the content evaluated, or which are only able to be observed in aggregate over some wider set of recommendations. These challenges can be mitigated by broadening subjective evaluations to a more representative sample of the public, but that may raise other challenges around the costs of running these evaluations at that scale.

More specific metrics

In a related way, evaluation metrics could be improved by greater specificity and explicitness about what concept the metric is trying to measure and therefore explicitness about how different interpretations of the same high-level concept are weighted.[165] In particular, public service media organisations could be more explicit about the kind of diversity they want to optimise, e.g. unique content viewed, the balance of categories viewed or the representation of demographics and viewpoints across recommendations, and whether they care about each individual's exposure or exposure across all users.

---

165 van Es, K. F. (2017). 'An Impending Crisis of Imagination : Data-Driven Personalization in Public Service Broadcasters'. *Media@LSE*. Available at: https://dspace.library.uu.nl/handle/1874/358206

Longer-term metrics

Another issue identified is that most metrics used in the evaluation of recommendation systems, within public service media and beyond, are short-term metrics, measured in days or weeks, rather than years. Yet at least some of the goals of stakeholders will be longer-term than the metrics used to approximate them. Users may be interested in both immediate satisfaction and in discovering new content so they continue to be informed and entertained in the future. Businesses may both be trying to maximise quarterly profits and also trying to retain users into the future to maximise profits in the quarters to come.

Short-term metrics are not entirely ineffective at predicting long-term outcomes. Better outcomes right now could mean better outcomes months or years down the road, so long as the context the recommendation system is operating in stays relatively stable and the recommendation system itself doesn't change user behaviour in ways that lead to poorer long-term outcomes.

By definition, long-term consequences take a longer time to occur, and thus there is a longer waiting period between a change in the recommendation system and the resulting change in outcome. A longer period between action and evaluation also means a greater number of confounding variables which make it more challenging to assess the causal link between the change in the system and the change in outcomes.

Dietmar Jannach, Professor at the University of Klagenfurt, highlighted this was a problem across academic and industry evaluations, and that 'when Netflix changes the algorithms, they measure, let's see, six weeks, two months to try out different things in parallel and look what happens. I'm not sure they know what happens in the long run.'[166]

---

166   Interview with Dietmar Jannach, Professor, University of Klagenfurt (2021).

Methods

Simulation-based evaluation

One possible method to estimate long-term metrics is to use **simulation-based offline evaluation approaches**. In this approach, the developers use a virtual environment with a set of content which can be recommended and a user model which simulates the expected preferences of users based on parameters selected by the developers (which could include interests, demographics, time already spent on the product, previous interactions with the product etc.).[167] This recommendation system then makes recommendations to the user model, which generates a simulated response to that recommendation. The user model can also update its preferences in response to the recommendations it has received, e.g. a user might become more or less interested in a particular category of content, and model the simulated users' overall satisfaction with the recommendations over time.

This provides some indication of how the dynamics of the recommendation system and changes to it might play out over a long period of time. It can evaluate how users respond to a series of recommendations over time and therefore whether a recommendation system could lead to audience satisfaction or diverse content exposure over a period longer than a single recommendation or user session. However, this approach still has many of the limitations of other kinds of offline evaluation. Historical user interaction data is still required to model the preferences of users, and that data is not neutral because it is itself the product of interaction with the previous system, including any previous recommendation system that was in place.

The user model is also only based on data from previous users, which might not generalise well to new users. Given that many of these recommendation systems are put in place to reach new audiences, specifically younger and more diverse audiences than those who currently use the service, the simulation-based evaluation might lead to unintentionally underserving those audiences and overfitting to existing user preferences.

---

167 Ie, E., Hsu, C., Mladenov, M. et al. (2019). 'RecSim: A Configurable Simulation Platform for Recommender Systems'. *arXiv*. Available at: https://doi.org/10.48550/arXiv.1909.04847

Furthermore, the simulation can only model the impact of parameters coded into it by the developers. The simulation only reflects the world as a developer understands it, and may not reflect the real considerations users take into account in interacting with recommendation systems, nor the influences on user behaviour beyond the product.

This means that if there are unexpected shocks, exogenous to the recommendation system, that change user interaction behaviour to a significant degree, then the simulation will not take those factors into account. For example, a simulation of a news recommendation system's behaviour in December 2019 would not be a good source of truth for a recommendation system in operation during the COVID-19 pandemic. The further the simulation tries to look ahead at outcomes, the more vulnerable it will be to changes in the environment that may invalidate its results.

User panels and retrospective feedback

After deployment, asking audiences for informed and retrospective feedback on their recommendations is a promising method for short-term and long-term recommendation system evaluation.[168] This could involve asking the users to review, rate and provide feedback on a subsection of the recommendations they received over the previous month, in a similar manner to the subjective evaluations undertaken by the BBC Datalab. This would provide development and product teams with much more informative feedback than through A/B testing.

This could be particularly effective in the form of a representative longitudinal user panel which returns to the same audience members at regular intervals to get their detailed feedback on recommendations.[169] Participants in these panels should be compensated for their participations to recognise the contribution they are making to the improvement of the system and ensure long-term retention of participants. This would allow development and product teams to gauge how audience responses change over time, by seeing how they react

168 Stray, J., Adler, S. and Hadfield-Menell, D. (2020), 'What are you optimizing for? Aligning Recommender Systems with Human Values',
      pp. 4–5. *Participatory Approaches to Machine Learning ICML 2020 Workshop (July 17).* Available at:
      https://participatoryml.github.io/papers/2020/42.pdf

169 Stray, J. (2021). 'Beyond Engagement: Aligning Algorithmic Recommendations With Prosocial Goals'. *Partnership on AI.* Available at:
      https://www.partnershiponai.org/beyond-engagement-aligning-algorithmic-recommendations-with-prosocial-goals/

to the same recommendations months later, to understand how their opinions on that recommendation may have changed over time, including in response to changes to the underlying system over longer periods.

Case studies

Through two case studies, we examine how the differing prioritisation of values in different forms of public service media and the differing nature of the content itself manifests itself in different approaches to recommendation systems. We will focus on the use of recommendation systems across BBC News for news content, and BBC Sounds for audio-on-demand.

Case study 1: BBC News

Introduction

BBC News is the UK's dominant news provider and one of the world's most influential news organisations.[170] It reaches 57% of UK adults every week and 456 million globally. Its news websites are the most-visited English language news websites on the internet.[171] For most of the time that BBC News has had an online presence, it has not used any recommendation systems on its platforms.

In recent years, BBC News has taken a more experimental approach to recommendation systems, with a number of different systems for recommending news content developed, piloted and deployed across

---

170 This case study focuses on the parts of BBC News that function as a public service, rather than BBC Global News, the international commercial news division.

171 As of 2021, BBC News on TV and radio reaches 57% of UK adults every week and across all channels, BBC News globally reaches a weekly global audience of 456 million adults. See: BBC Media Centre. (2021). 'BBC on track to reach half a billion people globally ahead of its centenary in 2022'. Available at: https://www.bbc.co.uk/mediacentre/2021/bbc-reaches-record-global-audience; BBC News is equally influential globally within the domain of digital news. By one measure, the BBC News and BBC World News websites combined are the most-visited English-language news websites, receiving three to four times the website traffic of the *New York Times*, *Daily Mail*, or *The Guardian*, see: Majid, A. (2021). 'Top 50 largest news websites in the world: Surge in traffic to Epoch Times and other ring-wing sites'. *Press Gazette*. Available at: https://pressgazette.co.uk/top-50-largest-news-websites-in-the-world-right-wing-outlets-see-biggest-growth/; As of 2021, BBC News Online reaches 45% of UK adults every week, approximately triple the reach of its nearest competitors: *The Guardian* (17%), *Sky News Online* (14%) and the *MailOnline* (14%). Estimates of UK reach are based on a sample 2029 adults surveyed by YouGov (and their partners) using an online questionnaire at the end of January and beginning of February 2021. See: Reuters Institute for Institute for the Study of Journalism. *Reuters Institute Digital News Report 2021*, 10th Edition, p. 62. Available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf

the organisation.[172]

Goal

For editorial teams, the goal of adding recommendation systems to BBC News was to augment editorial curation and make it easier to scale on a more personalised level. This addresses challenges relating to editors facing an 'information overload' of content to recommend. Additionally, product teams at BBC believed this feature would improve the discoverability of news content for different users.[173]

What did they build?

From around 2019, a team (which later become part of BBC Datalab) collaborated with a team building out the BBC News app to develop a content-to-content recommendation system. This focused on 'onward journeys' from news articles. Partway through each article the recommendation system generated a section that was titled 'You might be interested in' (in the language relevant to that news website) that listed four recommended articles.[174]

172   The team initially developed an experimental recommendation system for BBC Mundo, the BBC World Service's Spanish-language news website. See: Panteli, M., Piscopo, A., Harland, A., Tutcher, J. and Moss, F. M. (2019). 'Recommendation systems for news articles at the BBC', p. 4. *CEUR Workshop Proceedings*. Available at: http://ceur-ws.org/Vol-2554/paper_07.pdf; These are also live on BBC World Service websites in Russian, Hindi and Arabic and in beta on the BBC News App. See: Piscopo, A. (2021). 'Building public service recommenders: Logbook of a journey' [presentation recording]. *The Academic Fringe Festival*. Available at: https://www.youtube.com/watch?v=Q2EYAxX5Pnk; Al-Chueyr Martins, T. (2019). 'Responsible Machine Learning at the BBC' [presentation]. Available at: https://www.slideshare.net/alchueyr/responsible-machine-learning-at-the-bbc-194466504

173   See: Panteli, M., Piscopo, A., Harland, A., Tutcher, J. and Moss, F. M. (2019). 'Recommendation systems for news articlesat the BBC', p. 4. *CEUR Workshop Proceedings*. Available at: http://ceur-ws.org/Vol-2554/paper_07.pdf

174   Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

Figure 2: BBC News 'You might be interested in' section (image courtesy of the BBC)

The recommendation system is combined with a set of business rules which constrain the set of articles that the system recommends content from. The rules aim to ensure 'sufficient quality, breadth, and depth' in the recommendations.[175] For example, these included:

- recency, e.g. only selecting content from the past few weeks
- unwanted content, e.g. content in the wrong language
- contempt of court
- elections
- children-safe content.

In an earlier project, this team had developed an experimental recommendation system for BBC Mundo, the BBC World Service's Spanish-language news website.[176] Similar recommendation systems are also live on BBC World Service websites in Russian, Hindi and Arabic and in beta on the BBC News App.[177]

175 Piscopo, A. (2021). 'Building public service recommenders: Logbook of a journey' [presentation recording]. *The Academic Fringe Festival*. Available at: https://www.youtube.com/watch?v=Q2EYAxX5Pnk

176 Panteli, M., Piscopo, A., Harland, A., Tutcher, J. and Moss, F. M. (2019). 'Recommendation systems for news articles at the BBC', p. 4. *CEUR Workshop Proceedings*. Available at: http://ceur-ws.org/Vol-2554/paper_07.pdf

177 Piscopo, A. (2021). 'Building public service recommenders: Logbook of a journey' [presentation recording]. *The Academic Fringe Festival*. Available at: https://www.youtube.com/watch?v=Q2EYAxX5Pnk; Al-Chueyr Martins, T. (2019). 'Responsible Machine Learning at the BBC' [presentation]. Available at: https://www.slideshare.net/alchueyr/responsible-machine-learning-at-the-bbc-194466504

Figure 3: BBC Mundo recommendation system (image courtesy of the BBC)

Figure 4: Recommendation system on BBC World Service website in Hindi (image courtesy of the BBC)

ये भी पढ़ें

टी20 विश्व कप: बांग्लादेश को हरा सेमीफ़ाइनल की राह आसान कर सकता है भारत

विराट कोहली : क्या क्रिकेट पिच पर ये उनका 'दूसरा जन्म' है?

दक्षिण अफ़्रीका से भिड़ने से पहले भारत की बढ़ी चिंता, पाकिस्तान वाला डर

टी-20 वर्ल्ड कप: 'काग़ज़ी शेर' भारतीय क्रिकेट टीम बड़ी प्रतियोगिता में ढेर क्यों हो जाती है?

Criteria (and how they relate to public service values)

The BBC News team eventually settled on a content-to-content recommendation system using a model (called 'tf-idf') that encoded article data (like text) and metadata (like the categorical tags that editorial teams gave the article) into vectors. Once articles were represented as vectors, additional metrics could be applied to measure the similarity between them. This enabled the ability to penalise more popular content.[178]

The business rules the BBC used sought to ensure 'sufficient quality, breadth, and depth' in the recommendations, which aligns with the BBC's values around universality and excellence.[179]

There was also an emphasis on the recommendation system needing to be easy to understand and explain. This can be attributed to BBC News being more risk-averse than other parts of the organisation.[180] Given the

---

178   Crooks, M. (2019). 'A Personalised Recommender from the BBC'. *BBC Data Science*. Available at: https://medium.com/bbc-data-science/a-personalised-recommender-from-the-bbc-237400178494

179   Piscopo, A. (2021). 'Building public service recommenders: Logbook of a journey' [presentation recording]. *The Academic Fringe Festival*. Available at: https://www.youtube.com/watch?v=Q2EYAxX5Pnk

180   Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

BBC's mandate to be a 'provider of accurate and unbiased information' and BBC News that staff themselves identify as 'the product that likely contributes most to its reputation as a trustworthy and authoritative media outlet'.[181] It is unsurprising they would want to pre-empt any accusations of bias for any automated news recommendation system, by making it understandable to audiences.

Evaluation

The Datalab team experimented with a number of approaches using a combination of content and user interaction data.

Initially, they found that a content-to-content approach to item recommendations was more suited to the editorial requirements for the product, and user interaction data was therefore less relevant to the evaluation of the recommender, prompting a shift to a different approach.

As they began to compare different content-to-content approaches, they found that performance in quantitative metrics often didn't match the editorial teams priorities, and it was difficult to operationalise editorial judgement of public service value into metrics. As Alessandro Piscopo notes: 'We did notice that in some cases, one of the recommender prototypes was going higher in some metrics and went to editorial and [they would] say well we just didn't like it.' And, 'Sometimes it was just comments from editorial world, we want to see more depth. We want to see more breadth. Then you have to interpret what that means.'[182]

The Datalab team chose to take a subjective evaluation-first approach, whereby editors would directly compare and comment on the output of two recommendation systems. This approach allowed them to capture editorial preferences more directly and manually work towards meeting those preferences.

---

181   Panteli, M., Piscopo, A., Harland, A., Tutcher, J. and Moss, F. M. (2019). 'Recommendation systems for news articles at the BBC', p. 4. *CEUR Workshop Proceedings*. Available at: http://ceur-ws.org/Vol-2554/paper_07.pdf

182   Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

However, the team acknowledged its limitations upfront, particularly in terms of scale.[183] They tried to pick articles that would bring up the most challenging cases. However, editorial teams only have so much capacity to judge recommendations, and thus would struggle to assess every edge case or judge every recommendation. This issue would be even more acute if in a future recommendation system, every article's associated recommendations changed depending on the demographics of the audience member viewing it.

By May 2020, editorial had given sign-off and the recommendation system was deemed ready for initial deployment.[184] During deployment, the team took a 'failsafe approach' with weekly monitoring of the live version of the recommendation system by editorial staff, alongside A/B testing measuring the click-through rate on the recommended articles. However, members of the development team noted it was only a rough proxy for user satisfaction and were working to go beyond click-through rate.

Case study 2: BBC Sounds

Introduction

BBC Sounds is the BBC's audio streaming and download service for live radio, music, audio-on-demand and podcasts,[185] replacing the BBC's previous live and catch-up audio service, iPlayer Radio.[186] A key difference between BBC Sounds and iPlayer Radio is that BBC Sounds was built with personalisation and recommendation as a core

183   Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

184   Piscopo, A. (2021). 'Building public service recommenders: Logbook of a journey' [presentation recording]. *The Academic Fringe Festival.* Available at: https://www.youtube.com/watch?v=Q2EYAxX5Pnk

185   BBC. 'What is BBC Sounds?'. Available at: https://www.bbc.co.uk/contact/questions/help-using-bbc-services/what-is-sounds

186   The BBC Sounds website replaced the iPlayer Radio website in October 2018; the BBC Sounds app was launched in beta in the United Kingdom in June 2018 and made available internationally in September 2020, with the iPlayer Radio app decommissioned for the United Kingdom in September 2019 and internationally in November 2020. See: BBC. (2018). 'The next major update for BBC Sounds' Available at: https://www.bbc.co.uk/blogs/aboutthebbc/entries/03e55526-e7b4-45de-b6f1-122697e129d9; BBC. (2018). 'Introducing the first version of BBC Sounds', Available at: https://www.bbc.co.uk/blogs/aboutthebbc/entries/bde59828-90ea-46ac-be5b-6926a07d93fb; BBC. (2020). 'An international update on BBC Sounds and BBC iPlayer Radio'. Available at: https://www.bbc.co.uk/blogs/internet/entries/166dfcba-54ec-4a44-b550-385c2076b36b; BBC Sounds. 'Why has the BBC closed the iPlayer Radio app?'. Available at: https://www.bbc.co.uk/sounds/help/questions/recent-changes-to-bbc-sounds/iplayer-radio-message

component of the product, rather than as a radio catch-up service.[187]

Goal

The goals of BBC Sounds as a product team are:

- increase the audience size of BBC Sounds' digital products
- increase the demographic breadth of consumption across BBC Sounds' products, especially among the young[188]
- convert 'lighter users' who only engage a certain number of times a week into regular users
- enable users to more easily discover content from the more than 50 hours of new audio produced by the BBC on an hourly basis.

Product

BBC Sounds initially used an outsourced recommendation system from a third-party provider. Having knowledge about the inner working of the recommendation systems and the ability to quickly iterate were seen as valuable by the development team, as it proved challenging to request changes to the external provider. The BBC decided it wanted to own the technology and the experience as a whole, and believed they could achieve better value-for-money for TV License-payers by bringing the system in-house. So the BBC Datalab developed a hybrid recommendation system named Xantus for BBC Sounds.

BBC Sounds use a factorisation machine approach, which is a mixture of content matching and collaborative filtering. This uses your listening history, metadata about the content, and other users' listening history

---

187   In May 2019, six months after the launch of BBC Sounds, James Purnell, then Director of Radio & Education at the BBC, said that 'The [BBC Sounds] app, for instance, is built for personalisation, but is not yet fully personalised. This means that right now a user sees programmes that have not been curated for them. That is changing, as of this month in fact. By the autumn, Sounds will be highly personalised.' See: BBC Media Centre. (2019). *Changing to stay the same - Speech by James Purnell, Director, Radio & Education, at the Radio Festival 2019 in London.* Available at:
https://www.bbc.co.uk/mediacentre/speeches/2019/bbc.com/mediacentre/speeches/2019/james-purnell-radio-festival/

188   According to David Jones (Executive Product Manager, BBC Sounds, interviewed in 2021), his top-line KPI is to reach 900,000 members of the British population who are under 35 by March 2022. These numbers are determined centrally by BBC senior managers based on the BBC's Service Licence for BBC Online and Red Button. See: BBC Trust. (2016). *BBC Online and Red Button Service Licence.* Available at:
http://downloads.bbc.co.uk/bbctrust/assets/files/pdf/regulatory_framework/service_licences/online/2016/online_red_button_may16.pdf
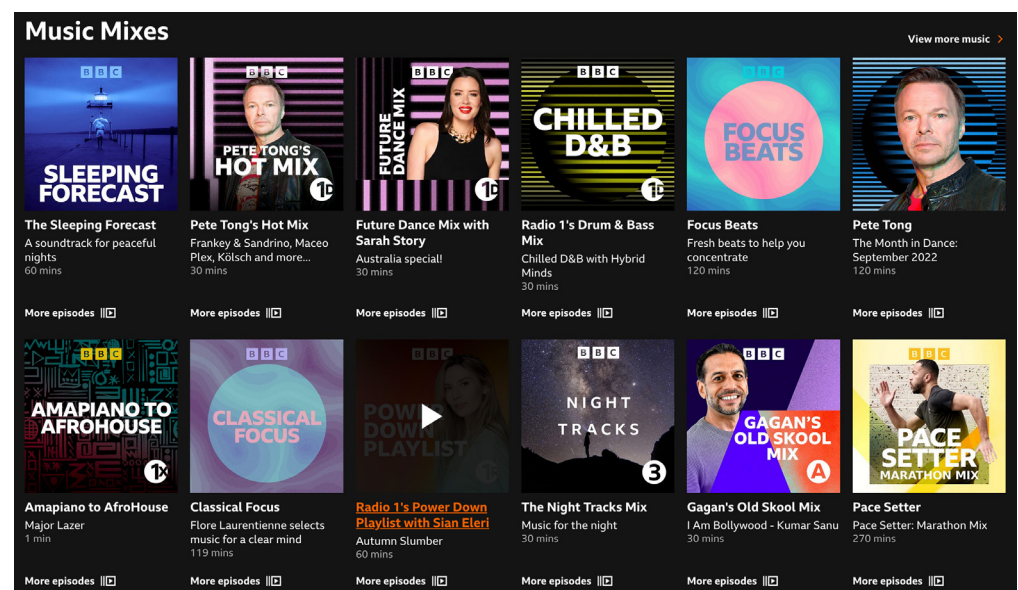
to make recommendations in two ways:

1.   It recommends items that have similar metadata to items you have already listened to.

2.   It recommends items that have been listened to by people with otherwise similar listening histories.

Figure 5: BBC Sounds' 'Recommended For You' section (image courtesy of the BBC)
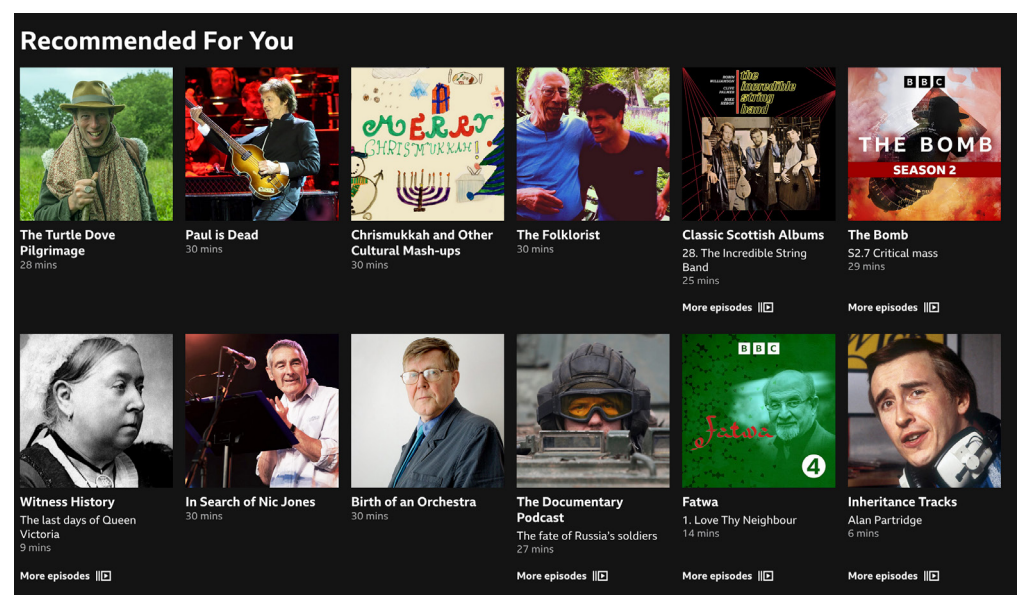
Figure 6: 'Music Mixes' on BBC Sounds (image courtesy of the BBC)

Criteria (and how they relate to public service media values)

On top of this factorisation machine approach are a number of business rules. Some rules apply equally across all users and constrain the set of content that the system recommends content from, e.g. only selecting content from the past few weeks. Other rules apply after individual user recommendations have been generated and filter the recommendations based on specific information about the user, e.g. not recommending content the user has already consumed.

As of summer 2021, the business rules used in the BBC Sounds' Xantus recommendation system were:[189]

Non-personalised business rules	Personalised business rules
Recency	Already seen items
Availability	Local radio (if not consumed previously)
Excluded 'master brands', e.g. particular radio channels	Specific language (if not consumed previously)
Excluded genres	Episode picking from a series
	Diversification (1 episode per brand/series)

Governance

Editorial and others help define the business rules for Sounds.[190] The product team adopted the business rules from the incumbent system and then checked whether they made sense in the context of the new system. They constantly review the business rules. Kate Goddard, Senior Product Manager, BBC Datalab, noted that:

'Making sure you are involving [editorial values] at every stage and making sure there is strong collaboration between data scientists in order to define business rules to make sure we can find good items.

---

189 Note that the business rules are subject to change, and so the rules given here are intended to be an indicative example only, representing a snapshot of practice at one point in time. See: Al-Chueyr Martins, T. (2021). 'From an idea to production: the journey of a recommendation engine' [presentation recording]. *MLOps London*. Available at: https://www.youtube.com/watch?v=dFXKJZNVgw4

190 Interview with Kate Goddard, Senior Product Manager, BBC Datalab (2021).

For instance with BBC Sounds you wouldn't want to be recommending news content to people that's more than a day or two old and that would be an editorial decision along with UX research and data. So, it's a combination of optimizing for engagement while making sure you are working collaboratively with editorial to make sure you have the right business rules in there.'

Evaluation

To decide whether to progress further with the prototype, the team decided to use a process of subjective evaluation. The Datalab team showed recommendations generated by both the new factorisation machine recommendation system head-to-head with the existing external provider's recommendations and got feedback from the editors on which of the two they liked.[191] The factorisation machine recommendation system was preferred by the editors and so was deployed into the live environment.

After deployment, UX testing, qualitative feedback and A/B testing were used to fine-tune the system. In their initial A/B tests, they were optimising for engagement, looking at click-throughs, play throughs and play completes. In these tests, they were able to achieve:[192]

• 59% increase in interactions in the 'Recommended for You' rail
• 103% increase in interactions for under-35s.

---

191   Interview with Alessandro Piscopo, Principal Data Scientist, BBC Datalab (2021).

192   Al-Chueyr Martins, T. (2021). 'From an idea to production: the journey of a recommendation engine' [presentation recording]. *MLOps London*. Available at: https://www.youtube.com/watch?v=dFXKJZNVgw4

Outstanding questions and areas for further research and experimentation

Through this research we have built up an understanding of the use of recommendation systems in public service media in the BBC and Europe, as well as the opportunities and challenges that arise. This section offers recommendations to address some of the issues that have been raised and indicate areas beyond the scope of this project that merit further research. These recommendations are directed at the research community, including funders, regulators and public service media organisations themselves.

There is an opportunity for public service media to define a new, responsible approach to the development of recommendation systems that work to the benefit of society as a whole and offer an alternative to the paradigm established by big technology platforms. Some initiatives that are already underway could underpin this, such as the BBC's Databox project with the University of Nottingham and subsequent work on developing personal data stores.[193] These personal data stores primarily aim to address issues around data ownership and portability, but could also act as a foundation for more holistic recommendations across platforms and greater user control over the data used in recommending them content.

But in making recommendations to public service media we recognise the pressures they face. In the course of this project, a real-terms cut to BBC funding has been announced and the corporation has said it will have to reduce the services it offers in response.[194] We acknowledge that, in the absence of new resources and faced with the reality of declining

193 Sharp, E. (2021). 'Personal data stores: building and trialling trusted data services'. *BBC Research & Development*. Available at: https://www.bbc.co.uk/rd/blog/2021-09-personal-data-store-research; Leonard, M. and Thompson, B. (2020), 'Putting audience data at the heart of the BBC'. *BBC Research & Development*. Available at: https://www.bbc.co.uk/rd/blog/2020-09-personal-data-store-privacy-services

194 Hansard – Volume 707: debated on Monday 17 January 2022. 'BBC Funding'. *UK Parliament*. Available at: https://hansard.parliament.uk//commons/2022-01-17/debates/7E590668-43C9-43D8-9C49-9D29B8530977/BBCFunding

budgets, public service media organisations would have to cut other activities to carry out our suggestions.

We therefore encourage both funders and regulators to support organisations to engage in public service innovation as they further explore the use of recommendation systems. Historically the BBC has set a precedent for using technology to serve the public good, and in doing so brought soft power benefits to the UK. As the UK implements its AI strategy, it should build on this strong track record and comparative advantage and invest in the research and implementation of responsible recommendation systems.

1. Define public service value for the digital age

Recommendation systems are designed to optimise against specific objectives. However, the development and implementation of recommendation systems is happening at a time when the concept of public service value and the role of public service media organisations in the wider media landscape is rapidly changing.

Although we make specific suggestions for approaches to these systems, unless public service media organisations are clear about their own identities and purpose, it will be difficult for them to build effective recommendation systems. It is essential that public service media revisit their values in the digital age, and articulate their role in the contemporary media ecosystem.

In the UK, significant work has already been done by Ofcom as well as the Digital, Culture, Media and Sport Select Committee to identify the challenges public service media face and offer new approaches to regulation. Their recommendations must be implemented so that public service media can operate within a paradigm appropriate to the digital age and build systems that address a relevant mission.

2. Fund a public R&D hub for recommendation systems and responsible recommendation challenges

There is a real opportunity to create a hub for the research and development of recommendation systems that are not tied to industry goals. This is especially important as recommendation systems are

one of the prime use cases of behaviour modification technology, but research into it is impaired by lack of access to interventional data.[195]

Existing academic work on responsible recommendations could be brought together into a public research hub on responsible recommendation technology, with the BBC as an industry partner. It could involve developing and deploying methods for democratic oversight of the objectives of recommendation systems and the creation and maintenance of useful datasets for researchers outside of private companies.

We recommend that the strategy for using recommendation systems in public service media should be integrated within a broader vision to make this part of a publicly accountable infrastructure for social scientific research.

Therefore, as part of UKRI's National AI Research and Innovation (R&I) Programme, set out in the UK AI Strategy, it should fund the development of a public research hub on recommendation technology. This programme could also connect with the European Broadcasting Union's PEACH project, which has similar goals and aims.

Furthermore, one of the programme's aims is to create challenge-driven AI research and innovation programmes for key UK priorities. The arrival of Netflix in 2006 spurred the development of today's recommendation systems. The UK could create new challenges to spur the development of responsible recommendation system approaches encouraging a better information environment. For example, the hub could release a dataset and benchmark for a challenge on generating automatic labels for a dataset of news items.

3. Publish research into audience expectations of personalisation

There was a striking consensus in our interviews with public service media teams working on recommendation systems that personalisation

195   Greene, T., Martens, D. and Shmueli, G. (2022). 'Barriers to academic data science research in the new realm of algorithmic behaviour modification by digital platforms'. *Nature Machine Intelligence*, 4, pp.323–330. Available at: https://www.nature.com/articles/s42256-022-00475-7

was both wanted and expected by the audience. However, we were offered little evidence to support this belief. Research in this area is essential for a number of reasons.

1. Public service media exist to serve the public. They must not assume they are acting in the public interest without any evidence of their audience's views towards recommendation systems.

2. The adoption of recommendation systems without evidence that they are either wanted or needed by the public raises the risk that public service media are blindly following a precedent set by commercial competitors, rather than defining a paradigm aligned to their own missions.

3. Public service media have limited resources and multiple demands. It is not strategic to invest heavily in the development and implementation of these systems without an evidence base to support their added value.

If research into user expectations of recommendation systems does exist, the BBC should strive to make this public.

4. Communicate and be transparent with audiences

Although most public service media organisations profess a commitment to transparency about their use of recommendation systems, in practice there is limited effective communication with their audiences about where and how recommendation systems are being used.

What communication there is tends to adopt the language of commercial services, for example talking about 'relevance'. In our interviews, we found that within teams there was no clear responsibility for audience communication. Staff often assumed that few people would want to know more, and that any information provided would only be accessed by a niche group of users and researchers.

However, we argue that public service organisations have a responsibility to explain their practices clearly and accessibly and to put their values of transparency into practice. This should not only help retain public trust at a time when scandals from big technology companies have understandably made people view algorithmic systems with suspicion,

but also develop a new, public service narrative around the use of these technologies.

Part of this task is to understand what a meaningful explanation of a recommendation system looks like. Describing the inner workings of algorithmic decision-making is not only unfeasible but probably unhelpful. However, they can educate audiences about the interactive nature of recommendation systems. They can make salient the idea that when consuming content through a recommendation system, they are in effect 'voting with their attention'. Their viewing behaviour is something private, but at the same time affects what the system learns and what others will view.

Public service media should invest time and research into understanding how to usefully and honestly articulate their use of recommendation systems in ways that are meaningful to their audiences.

This communication must not be one-way. There must be opportunities for audience members to give feedback and interrogate the use of the systems, and raise concerns where things have gone wrong.

5. Balance user control with convenience

However, transparency alone is not enough. Giving users agency over the recommendations they see is an important part of responsible recommendation. Simply giving users direct control over the recommendation system is an obvious and important first step, but it is not a universal solution.

Some interviewees pointed to evidence that the majority of users do not choose to use these controls and instead opt for the default setting. But there is also evidence that younger users are beginning to use a variety of accounts, browsers and devices, with different privacy settings and aimed at 'training' the recommendation algorithm to serve different purposes.

Many public service media staff we spoke with described providing this level of control. Some challenges that were identified include the difficulty of measuring how well the recommendations meet specific targets, as well as risks relating to the potential degradation of the user experience.

Firstly, some of our interviewees noted how it would be more difficult to measure how well the recommendation system is performing on dimensions such as diversity of exposure, if individual users were accessing recommendations through multiple accounts. Secondly, it was highlighted how recommendation systems are trained on user behavioural data, and therefore giving more latitude to users to intentionally influence the recommendations may give rise to negative dynamics that degrade the overall experience for all users over the long run, or even expose the system to hostile manipulation attempts.

While these are valid concerns, we believe that there is some space for experimentation, between giving users no control and too much control. For example, users could be allowed to have different linked profiles, and key metrics could be adjusted to take into account the content that is accessed across these profiles. Users could be more explicitly shown how to interact with the system to obtain different styles of recommendations, making it easy to maintain different 'internet personas'. Some form of ongoing monitoring for detecting adversarial attempts at influencing recommendation choices could also be explored. We encourage the BBC to experiment with these practices and publish research on their findings.

Another trial worth exploring is allowing 'joint' user recommendation profiles, where the recommendations are made based on multiple individuals' aggregated interaction history and preferences, such as a couple, a group of friends or a whole community. This would allow users to create their own communities and 'opt-in' to who and what influenced their recommendations in an intuitive way. This could enabled by the kind of personal data stores being explored by the BBC and Belgian VRT.[196]

There are multiple interesting versions of this approach. In one version, you would see recommendations 'meant' for others and know it was a recommendation based on their preferences. In another version, users would simply be exposed to a set of unmarked recommendations based on all their combined preferences.

Another potential approach to pilot would be to create different recommendation systems that coexist and allow users to choose which

196 Sharp, E. (2021). 'Personal data stores: building and trialling trusted data services'. *BBC Research & Development*. Available at: https://www.bbc.co.uk/rd/blog/2021-09-personal-data-store-research

they want to use or offer different ones at different times of day or when significant events happen (e.g. switching to a different recommendation system during the run up to an election or overriding them with breaking news). Such an approach might offer a chance to invite audiences to play a more active part in the formulation of recommendations, and open up opportunities for experimentation, which would need to be balanced against the additional operational costs that would be introduced.

6. Expand public participation

Beyond transparency or individual user choice and control over the parameters of the recommendation systems already deployed, users and wider society could also have greater input during the initial design of the recommendation systems and in the subsequent evaluations and iterations.

This is particularly salient for public service media organisations, as unlike private companies, which are primarily accountable to their customers and shareholders, public service media organisations see themselves as having a universal obligation to wider society. Therefore, even those who are not direct consumers of content should have a say in how public service media recommendations are shaped.

User panels

One approach to this, suggested by Jonathan Stray, is to create user panels that provide informed, retrospective feedback about live recommendation systems.[197] These would involve paying users for detailed, longitudinal data about their experiences with the recommendation system.

This could involve daily questions about their satisfaction with their recommendations, or monthly reviews where users are shown a summary of their recommendations and interaction with them. They could be asked how happy they are with the recommendations, how well are their interests served and how informed they feel.

197 Stray, J. (2021). 'Beyond Engagement: Aligning Algorithmic Recommendations With Prosocial Goals'. *Partnership on AI*. Available at: https://www.partnershiponai.org/beyond-engagement-aligning-algorithmic-recommendations-with-prosocial-goals/

This approach could provide new, richer and more detailed metrics for developers to optimise the recommendation systems against, which would potentially be more aligned with the interests of the audience. It might also open up the ability to try new approaches to recommendation, such as reinforcement learning techniques that optimise for positive responses to daily and monthly surveys.

Co-design

A more radical approach would be to involve audience communities directly in the design of the recommendation system. This could involve bringing together representative groups of citizens, analogous to citizens' assemblies, which have direct input and oversight of the creation of public service media recommendation systems, creating a third core pillar in the design process, alongside editorial teams and developer teams. This is an approach that has been proposed by the Media Reform Coalition Manifesto for a People's Media.[198]

These would allow citizens to ask questions of the editors and developers about how the system is intended to work, what kinds of data inform those systems and about what alternative approaches exist (including not using recommendation systems at all). These groups could then set out their requirements for the system and iteratively provide feedback on versions of the system as its developed, in the same way that editorial teams have, for example, by providing qualitative feedback on recommendations provided by different systems.

7. Standardise metadata

Each public service media organisation should have a central function that standardises the format, creation and maintenance of metadata across the organisation.

Inconsistent, poor quality metadata was consistently highlighted as a barrier to developing recommendation systems in public service media, particularly in developing more novel approaches that go beyond user

198   Grayson, D. (2021). *Manifesto for a People's Media*. Media Reform Coalition. Available at: https://drive.google.com/file/u/1/d/1_6GeXiDR3DGh1sYjFI_hbgV9HfLWzhPi/view?usp=embed_facebook

engagement and try to create diverse feeds of recommendations.

Institutionalising the collection of metadata and making access to it more transparent across each individual organisation is an important investment in public service media's future capabilities.

We also think it's worth exploring how much metadata can be standardised across European media organisations. The European Broadcasting Union (EBU)'s 'A European Perspective' project is already trialling bringing together content from across different European public service media organisations onto a single platform, underpinned by the EBU's PEACH system for recommendations and the EuroVOX toolkit for automated language services. Further cross-border collaboration could be enabled by sharing best practices among member organisations.

8. Create shared recommendation system resources

Some public service media organisations have found it valuable to have access to recommendations-as-a-service provided by the European Broadcasting Union (EBU) through their PEACH platform. This reduces the upfront investment required to start using the recommendation system and provides a template for recommendations that have already been tested and improved upon by other public service media organisations.

One area identified as valuable for the future development of PEACH was greater flexibility and customisation. For example, some asked for the ability to incorporate different concepts of diversity into the system and control the relative weighting of diversity. Others would have found it valuable to be able to incorporate more information on the public service value of content into the recommendations directly.

We also heard from several interviewees that they would value a similar repository for evaluating recommendation systems on metrics valued by public service media, including libraries in common coding languages, e.g. Python, and a number of worked examples for measuring the quality of recommendations. The development of this could be led by the EBU or a single organisation like the BBC.

This would help systemise the quantifying of public service values and collate case studies of how values are quantified. This would be best as

an open-source repository that others outside of public service media could learn from and draw on. This would:

- lower costs and thus easier to justify investment
- reduce the technical burden, making it easier for newer and smaller teams to implement
- point to how they're used elsewhere, reducing the burden of proof and making the alternative approach appear less risky
- provide source of existing ideas, meaning the team have to spend less time either coming up with their own (which might be suboptimal and discover that for themselves) or spend time wading through the technical literature.

Future public service media recommendation systems projects, and responsible recommendation system development more broadly, could then more easily evaluate their system against more sophisticated metrics than just engagement.

9. Create and empower integrated teams

When developing and deploying recommendation systems, public service media organisations need to integrate editorial and development teams from the start. This ensures that the goals of the recommendation system are better aligned with the organisation's goals as a whole and ensures the systems augment and complement existing editorial expertise.

An approach that we have seen applied successfully is having two project leads, one with an editorial background and one with a technical development background, who are jointly responsible for the project.

Public service media organisations could also consider adopting a combined product and content team. This can ensure that both editorial and development staff have a shared language and common context, which can reduce the burden of communication and help staff feel like they have a common purpose rather than competition between the different teams.

Methodology

To investigate our research questions, we adopted two main methods:

1. Literature review
2. Semi-structured interviews

Our literature review surveyed current approaches to recommendation systems, the motivations and risks in using recommendation systems, and existing approaches and challenges in evaluating recommendation systems. We then focused in on reviewing existing public information on the operation of recommendation systems across European public service media, and the existing theorical work and case studies on the ethics implications of the use of those systems.

In order to situate the use of these systems, we also surveyed the history and context of public service media organisations, with a particular focus on previous technological innovations and attempts at measuring values.

We also undertook 29 semi-structured interviews with 8 current and 3 former BBC staff members, across engineering, product and editorial, 9 interviews with current and former staff from other public service media organisations and the European Broadcasting Union, and 9 further interviews with external experts from academia, civil society and regulators.

Partner information and acknowledgements

Bozen-Bolzano

- Greg Detre, Chief Product & Technology Officer, Filtered and former Chief Data Scientist, Channel 4

- Jannick Kirk Sørensen, Associate Professor in Digital Media, Aalborg University

- Jonas Schlatterbeck, Head of Content ARD Online & Leiter Programmplanung, ARD

- Jonathan Stray, Visiting Scholar, Berkeley Center for Human-Compatible AI

- Kate Goddard, Senior Product Manager, BBC Datalab

- Koen Muylaert, Head of Data Platform, VRT

- Matthias Thar, Bayerische Rundfunk

- Myrna McGregor, BBC Lead, Responsible AI+ML

- Natalie Fenton, Professor of Media and Communications, Goldsmiths, University of London

- Nic Newman, Senior Research Associate, Reuters Institute for the Study of Journalism

- Olle Zachrison, Deputy News Commissioner & Head of Digital News Strategy, Swedish Radio

- Sébastien Noir, Head of Software, Technology and Innovation, European Broadcasting Union and Dmytro Petruk, Developer, European Broadcasting Union

- Sophie Chalk, Policy Advisor, Voice of the Listener & Viewer

- Uli Köppen, Head of AI + Automation Lab, Co-Lead BR Data, Bayerische Rundfunk

Annex 1: How widely used are recommendation systems among public service media?

The European Broadcasting Union (EBU) is a membership association of public service media organisations.[199] Members may represent more than one organisation, for example the German ARD represents 11 regional and national broadcasters. Membership is restricted to public service media organisations in countries within the European Broadcasting Area, as defined by the International Telecommunication Union, or are members of the Council of Europe. There are 69 members representing 115 organizations in 56 countries, including 15 members and 16 organisations based in North Africa and Western Asia.

The EBU's Media Intelligence Service runs annual member surveys on a number of topics, including the use of recommendation systems. In 2020, only 46 members responded to the EBU's question on sign-in features, and the questions on personalised recommendations only represent at most the 32 members who reported having a sign-on feature. 39 to 42 organisations answered questions on which platform, if any, used personalised recommendations, and 22 organisations answered questions on which part of the platform the personalised recommendations were displayed.[200]

Thus the statistics may undercount the number of public service media organisations in Europe deploying recommendation systems if recommendation systems are being used by those who didn't respond to the survey or who don't use a sign-on feature, or who do use a sign-on feature but don't personalise their recommendations. The statistics may instead overcount the number of public service media organisations in Europe deploying recommendation systems if the members who responded to the question on sign-on features were disproportionately those outside of Europe.

199 European Broadcasting Union (EBU). (2021). 'Our Members'. Available at: https://www.ebu.ch/about/members
200 Correspondence with the EBU Media Intelligence Service, based on 2020 Members' data.

Based on our literature review and a scan of public service media organisation websites, it seems likely that this statistic is an undercount rather than overcount. However, it at least provides a baseline for the prevalence of recommendation systems among European public service media in 2020/2021.

However, estimates of prevalence don't tell us how widely or deeply recommendation systems are used by European public service media.

Annex 2: A brief history of the BBC's development and use of recommendation systems

The BBC has a history of developing and deploying recommendation systems going back at least as far as 2008 with their participation in the EU-funded MyMedia research project, which developed recommendation systems for multimedia content.[201] The BBC helped evaluate the MyMedia collaborative filtering recommendation system through a field trial on online catch-up services for TV and radio.[202]

The BBC continued to trial systems based on MyMedia into the early 2010s. This included:

- in 2011, field trials comparing matrix factorisation techniques against meta-data-based approaches for TV recommendations[203]

- from 2012 onwards, the creation and trialling of Sibyl, a prototype privacy-preserving recommendation system for TV and radio content, first just using programme metadata and then expanding to a hybrid system include collaborating filtering[204]

- in 2014, extensions of the MyMediaLite Recommendation System library with diversity metrics such as surprisal.

201 BBC Research & Development. 'MyMedia'. Available at: https://www.bbc.co.uk/rd/projects/mymedia; European Commission. (2022). 'Dynamic Personalization of Multimedia – MyMedia Project'. Available at: https://cordis.europa.eu/project/id/215006/results

202 Newell, C. and Knijnenburg, B. (2010). *Enhanced Internet A/V Content Field Trial Report*. BBC Research & Development. Available at: https://downloads.bbc.co.uk/rd/pubs/mymedia/mymedia-deliverable54.pdf

203 Nicolaou, A. (2011). 'Prototyping weeknotes #87'. *BBC R&D Blog*. Available at: https://www.bbc.co.uk/rd/blog/2011-11-prototyping-weeknotes-87; Miller, L. (2012). 'IRFS Weeknotes #111'. *BBC Research & Development*. Available at: https://www.bbc.co.uk/rd/blog/2012-06-irfs-weeknotes-111; Gregory-Clarke, R. (2012). 'Subjective Assessment of a User-controlled Interface for a TV Recommender System'. *European Conference on Interactive TV*; Gregory-Clarke, R. and Newell, C. (2012). 'TV Recommender System Field Trial Using Dynamic Collaborative Filtering'. *European Conference on Interactive TV*.

204 Newell, C. (2012). 'Sibyl Recommender Update'. *BBC Research & Development*. Available at: https://www.bbc.co.uk/blogs/researchanddevelopment/2012/11/sibyl-recommender-update.shtml; BBC Research & Development. 'Sibyl Recommender System'. Available at: https://www.bbc.co.uk/rd/projects/sibyl-recommender-system

A renewed interest in recommendation system use begins with the Editorial Algorithms project in 2014 and the myBBC initiative in 2015.[205] This asked audiences to register an account with the BBC so they could collect more information about the audience's preferences and consumption patterns to inform personalisation.[206]

In 2017, the BBC Datalab was established with the aim of helping audiences discover relevant content by bringing together data from across the BBC, augmented machine learning and editorial expertise.[207] It was envisioned as a central capability across whole BBC (TV, radio, news, web) which would build a data platform for other BBC teams that creates a consistent and relevant experiences for audiences across different products. In practice, this has meant collaborating with different product teams to develop recommendation systems.

Since 2018, the use of recommendation systems has become more integrated into operations across the BBC. This is in part due the creation of central recommendation systems development team in the form of BBC Datalab and, since 2020, the BBC bringing all its recommendation systems in-house.

205 BBC Research & Development. 'Editorial Algorithms'. Available at: https://www.bbc.co.uk/rd/projects/editorial-algorithms

206 Fearnley, P. (2015). '; Miller, L. (2016). 'IRFS Weeknotes #222'. *BBC Research & Development*. Available at: https://www.bbc.co.uk/rd/blog/2016-06-irfs-weeknotes-number-222; Thereaux, O. (2017). 'IRFS Weeknotes #244'. *BBC Research & Development*. Available at: https://www.bbc.co.uk/rd/blog/2017-05-irfs-weeknotes-number-244

207 BBC. (2019). 'Join the DataLab team at the BBC!' *BBC Careers*. Available at: https://careerssearch.bbc.co.uk/jobs/job/Join-the-DataLab-team-at-the-BBC/40012; BBC Datalab. 'Machine learning at the BBC'. Available at: https://datalab.rocks/

About the Ada Lovelace Institute

The Ada Lovelace Institute was established by the Nuffield Foundation in early 2018, in collaboration with the Alan Turing Institute, the Royal Society, the British Academy, the Royal Statistical Society, the Wellcome Trust, Luminate, techUK and the Nuffield Council on Bioethics.

The mission of the Ada Lovelace Institute is to ensure that data and AI work for people and society. We believe that a world where data and AI work for people and society is a world in which the opportunities, benefits and privileges generated by data and AI are justly and equitably distributed and experienced.

We recognise the power asymmetries that exist in ethical and legal debates around the development of data-driven technologies, and will represent people in those conversations. We focus not on the types of technologies we want to build, but on the types of societies we want to build.

Through research, policy and practice, we aim to ensure that the transformative power of data and AI is used and harnessed in ways that maximise social wellbeing and put technology at the service of humanity.

We are funded by the Nuffield Foundation, an independent charitable trust with a mission to advance social well-being. The Foundation funds research that informs social policy, primarily in education, welfare and justice. It also provides opportunities for young people to develop skills and confidence in STEM and research. In addition to the Ada Lovelace Institute, the Foundation is also the founder and co-funder of the Nuffield Council on Bioethics and the Nuffield Family Justice Observatory.

Find out more:

Website: Adalovelaceinstitute.org
Twitter: @AdaLovelaceInst
Email: hello@adalovelaceinstitute.org