

Rebuilding Public Trust in the New Zealand Digital Media Industry: Explainability, Data Ethics and Governance

Aditya Santosh Pawade, Indrapriya Kularatne, and Shafiq Alam

AIC OTAGO POLYTECHNIC

ABSTRACT

The rapid adoption of AI-driven recommendation systems in New Zealand's digital media industry has transformed how audiences can access news, while simultaneously contributing to declining public trust. This study examines how explainable artificial intelligence, supported by data ethics and governance frameworks, can be used to rebuild trust in algorithmically curated news environments. A qualitative, desk-based research design was employed, using a systematic literature review of 38 academic and policy sources selected from an initial pool of 60 articles. Thematic synthesis identified four interconnected problem areas: declining trust in news accuracy, algorithmic opacity resulting in limited user comprehension, risks to fairness and editorial independence, and unresolved privacy and accountability concerns. The findings demonstrated that trust in AI-driven media systems is conditional on transparency, fairness, user control, and responsible data governance. This study proposes an integrated framework for trustworthy AI that aligns explainability with user-centred design, editorial practices, and culturally grounded data governance, including Māori sovereignty principles. The research contributes to both theory and practice by offering context-specific recommendations for New Zealand media organisations to deliver trustworthy personalised news systems while maintaining their democratic role in society.

Keywords: Explainable AI; News Recommender Systems; Media Trust; Algorithmic Transparency; Data Governance.

INTRODUCTION

The New Zealand digital media industry is transforming exponentially, with news organisations becoming increasingly dependent on AI based recommendation systems in order to personalise content and keep audiences in competitive online spaces (Ministry for Culture and Heritage, 2024). This technological change is occurring simultaneously with a decreased trust in news, as well as an increasing anxiety about misinformation and what New Zealanders see, understand and participate in. The credibility of such systems is a primary managerial concern. This paper discusses the ways in which digital media organisations in New Zealand can restore people's trust in the use of algorithmic recommendation systems in their work, with transparency, fairness, privacy protection, and accountable governance acting as the main pillars of trustful AI in the news.

Explainable Artificial Intelligence (XAI) is a set of methods and design approaches that make algorithmic decision-making processes understandable and interpretable to human users, enabling transparency in how outputs are generated (Miller, 2019). Within digital media, XAI is increasingly applied in news recommendation systems to provide user-facing explanations such as why particular content is recommended based on behavioral patterns, preferences, or prior interactions. These recommendation systems play a critical role in managing information exposure, acting as gatekeepers of civic and political content in personalised news environments. However, in many cases, such recommendation processes remain opaque, limiting users' ability to understand, question, or trust algorithmic decisions,

thereby reinforcing concerns about bias, manipulation, and lack of accountability.

A critical gap therefore remains in defining how transparency, fairness, privacy, and governance can be operationalised collectively to rebuild trust in algorithmically curated news environments within New Zealand. While existing studies have examined aspects of recommender systems and algorithmic transparency, limited attention has been given to their application in addressing declining public trust in the news. This gap is particularly significant in New Zealand where evolving regulatory expectations and culturally grounded data governance principles, including Māori data sovereignty, shape public perceptions of responsible AI use. Therefore, this study aims to examine how explainable AI, supported by robust data ethics and governance mechanisms, can be implemented by New Zealand digital media organisations to restore and sustain public trust in AI-driven news recommendation systems.

LITERATURE REVIEW

The digital media environment in New Zealand is in the process of rapid technological and economic change with organisations such as New Zealand Media and Entertainment (NZME), Stuff, Radio New Zealand (RNZ), and Television New Zealand (TVNZ) all becoming more dependent on algorithmic news delivery to digital platforms (Ministry for Culture and Heritage, 2024). New Zealand has one of the highest levels of internet access in the world, with 96.2% of the population online and almost half of them spending over four hours daily online (DataReportal, 2025). Social media use is also widespread with 79.1% of New Zealanders having used at least one social media platform, supporting the dominance of algorithmic feeds in news exposure (Carney, 2025). New Zealand On Air (2024) reports that younger audiences are now dependent largely upon digital, on demand and mobile platforms and are moving toward more personalised recommendation systems as opposed to established broadcast schedules.

This transition is occurring in the media sector which is experiencing significant economic strain. The media and broadcasting industry employs 25,141 people (Ministry for Culture and Heritage, 2024). However declining advertising revenue, newsroom retrenchment, and regional media contraction have reduced their editorial capacity and increased their dependence on engagement driven algorithms to retain audiences (Radio New Zealand, 2024). Media industry report 2024 shows uneven distribution and shrinkage of regional news, highlighting the structural limitations that make algorithmic personalisation more attractive to media organisations (Ministry for Culture and Heritage, 2024). These commercial pressures, however, risk narrowing editorial independence and restricting content diversity (New Zealand On Air, 2024).

InternetNZ (2024) reports that 50% of New Zealanders spend four or more hours online each day, rising to 67% among 18 to 29-year-olds. This exposure means algorithmic systems hold growing influence over users' attention and trust. Public trust in New Zealand news media has fallen sharply. Myllylahti and Treadwell (2025) have reported a decreased trust in news from 53% in 2020 to 32% in 2025, showing the rising perceptions that news is biased, sensational, or incomplete, a pattern directly linked to declining trust in accuracy. Sensationalism and attention optimised headlines suggest to audiences that engagement outweighs factual integrity, intensifying distrust (Wong et al., 2023). Authority (2024) reinforces this view through complaint trends which identify accuracy, balance, and fairness concerns. Brand specific trust disparities further demonstrate the level of trust fragmentation across outlets (Myllylahti & Treadwell, 2025).

The decline in media trust mirrors a reduction in institutional trust in New Zealand. Stats New Zealand (2024) has recorded falling trust in government, public authorities and media, illustrating an environment of scepticism. InternetNZ (2024) additionally reports that 61% of New Zealanders are very or extremely concerned about misinformation and 65% about the security of personal data, and that only 25% consider they know a lot or a fair amount

about AI. These are separate survey items and are treated separately here. These concerns clearly demonstrate how algorithmic systems must operate within an already low trust environment (FigureNZ, 2023).

Despite public uncertainty, use of AI continues to grow. AI Forum NZ (2025) reports that 93% of AI using businesses have realised efficiency gains while 56% report measurable financial benefits. This suggests that AI is reshaping national economic and institutional processes, including media operations.

Public sector adoption of AI mirrors this trend. The inter-agency AI survey of 2025 recorded 272 AI use cases across 70 agencies, a significant increase from 108 use cases across 37 agencies in 2024 (Digital Govt NZ, 2025). This growth concerns public-sector adoption rather than news recommendation and is reported here as context for rising public familiarity with algorithmic decision-making rather than as evidence of recommender-system use in New Zealand media. These findings illustrate how rapid institutional normalisation of algorithmic decision making has been institutionalised and raise expectations that private media organisations must adopt similar transparency and governance practices.

Financial contraction continues to intensify pressures to automate content delivery. NZME reported a NZ\$16 million net loss in 2024 despite modest revenue increases (NZME, 2025). This fragile fiscal environment makes engagement driven algorithms appealing as cost effective mechanisms to retain audiences. However, such reliance risks weakening fairness, diversity, and editorial independence (New Zealand On Air, 2024).

The New Zealand policy environment sets clear expectations for trustworthy AI. The Privacy Act 2020 excludes a news entity from the definition of an agency to the extent that it is carrying on news activities, so the Act does not apply uniformly to everything a media company does. That exclusion does not extend to subscriptions, customer accounts, advertising, analytics, behavioral profiling or the personalisation of content delivery, in respect of which the information privacy principles apply in full. Commissioner (2023) recommends, as best practice rather than as a statutory requirement, that organisations conduct a privacy impact assessment, are transparent about when and why AI is used, engage with Māori about risks to the taonga of their information, and ensure human review before acting on AI outputs. The MBIE Responsible AI Guidance emphasises governance structures, senior accountability and meaningful user facing disclosure (MBIE, 2025b). The Algorithm Charter for New Zealand formalises transparency, human oversight, and explainability as public sector norms (Data.Govt.NZ, 2020). Although private media entities do not sign the Charter, it provides a baseline toward which public expectations are converging.

The Ministry of Business publication “Strategic framework: Better digital future for Aotearoa” (2022) reinforces cultural expectations that digital systems must be open, inclusive and trusted. These frameworks together imply that explainability is no longer a matter of choice, but the key to preserving legitimacy.

Recommendation systems are automated systems that select, rank and personalise news according to user behaviour, reading patterns and previous interactions. Because these systems effectively act as gatekeepers of information, audiences must trust that the ranking processes are fair, accurate and not manipulative. Research shows that when users cannot understand why content appears in their feed or when content appears sensational or biased, trust declines sharply (Blassnig et al., 2024; Wong et al., 2023). Global platforms have started to implement ‘Why am I seeing this’ explanations (Blassnig et al., 2024). No published source identified in this review documents equivalent public-facing mechanisms in New Zealand news services, which is an absence of evidence rather than evidence of absence. In the context of low trust which is marked by data concerns and declining editorial capacity, opaque recommendation systems can create significant managerial risk.

New Zealand digital media organisations increasingly rely on AI driven news recommendation systems to retain audiences but operate in low trust environments with a lack of transparency, explanation or privacy clarity. This lack

of explainability threatens to undermine public confidence further and thus undermines the credibility of those who work in the field such as editors, and therefore undermine the democratic role of the news media in New Zealand.

Algorithmic news recommendation is an established feature of digital news distribution internationally and operates in New Zealand within an environment marked by declining public trust, limited transparency, concerns about fairness and persistent privacy risks (InternetNZ, 2024; Myllylahti & Treadwell, 2025). These systems increasingly shape how New Zealanders encounter political and civic information, yet they do so with minimal public explanation (Blassnig et al., 2024; Wong et al., 2023). Based on the reviewed literature, the following four critical problem areas have been identified: (1) public trust and accuracy crisis, (2) opacity and comprehension gap, (3) risks to fairness and editorial independence, and (4) unresolved privacy and accountability challenges (Blassnig et al., 2024; Commissioner, 2023; Jin et al., 2023; Myllylahti & Treadwell, 2025).

Public Trust and Accuracy Crisis

Public trust in the New Zealand news media has declined sharply from 53% in 2020 to 32% in 2025, representing one of the steepest drops among comparable democracies (Myllylahti & Treadwell, 2025). This decline is driven by perceptions of bias, sensationalism, and weakening editorial standards, with accuracy, balance, and misleading headlines being the most common public complaints (Authority, 2024). Broader declines in institutional trust further intensify scepticism towards media systems (Stats New Zealand, 2024). In this context, algorithmic news recommendation systems amplify distrust, particularly when users do not understand how content is selected (Blassnig et al., 2024). Trust in data-driven media is conditional on perceived transparency, fairness, and safety (Wong et al., 2023), yet a high level of concern about misinformation (61%) deepens uncertainty (InternetNZ, 2024). Consequently, algorithmic personalisation operates within an already fragile trust environment, where perceived manipulation and sensationalism further erode public confidence in both news systems and organisations.

Algorithmic Opacity and User Comprehension Gap

The publicly available sector literature documents the shift of major New Zealand media organisations such as NZME, Stuff, RNZ and TVNZ to digital distribution, but does not establish the extent, design or governance of algorithmic recommendation within these organisations (Ministry for Culture and Heritage, 2024). No source identified in this review describes how such systems operate or are explained to audiences, and that absence of public information is itself relevant to the argument advanced here. This contrasts with global platforms such as YouTube and TikTok, where ‘*Why am I seeing this*’ features are becoming standard (Blassnig et al., 2024). Despite rising exposure to AI-mediated content, only 25% of New Zealanders report understanding AI, with most expressing concern rather than optimism (InternetNZ, 2024). Explainable AI (XAI) research highlights that trust depends on clear, relevant, and user-centered explanations (Alam et al., 2026; Miller, 2019; Tintarev & Masthoff, 2015). Narrative and example-based explanations are more effective than technical explanations in improving comprehension and trust (Liao & Varshney, 2021; Sunny, 2025). Users also expect transparency and control over recommendations (Mitova et al., 2024). In this context, persistent opacity reinforces suspicion, reduces perceived legitimacy, and further weakens trust in algorithmically curated news (Blassnig et al., 2024).

Perceived Bias, Filter Bubbles and Editorial Independence

New Zealand newsrooms face increasing commercial pressures, including layoffs, reduced investigative capacity, and declining advertising revenue, leading to greater reliance on engagement-driven algorithms (Ministry for Culture and Heritage, 2024; Zealand, 2024). Research highlights that such systems can reinforce filter bubbles, ideological bias, and influence editorial decision-making (Diakopoulos & Koliska, 2017). In New Zealand, audiences already perceive news as biased, with concerns about AI-generated content and underrepresentation of Māori and Pasifika communities (Hope, W., Baker, S., Daniels, R., Hoar, P., McEwan, R., Myllylahti, M., Mountfort, P., Middleton, A., Selman Julian, D., Treadwell, G. & Watts, T., 2025; Kantar, 2021). Younger audiences increasingly consume news through algorithmic feeds, intensifying these effects. Fairness research shows that recommender systems often overexpose popular content while marginalising minority perspectives unless explicitly calibrated for diversity (Biega et al., 2018; Jin et al., 2023). As a result, perceived bias and reduced diversity weaken trust, making algorithmic personalisation appear manipulative, and undermining confidence in news systems (Blassnig et al., 2024).

Privacy, Data Ethics and Accountability Risks

Personalisation in digital media relies on behavioral, device-level, and engagement data, yet 65% of New Zealanders are very or extremely concerned about the security of personal data and 61% about identity theft (InternetNZ, 2024). The Privacy Act 2020 sets requirements for transparency, purpose limitation and accountability in the handling of personal information, and the Office of the Commissioner (2023) sets out how these apply across the lifecycle of AI systems. Because personalisation and analytics fall outside the news-activity exclusion, these requirements reach precisely the data practices that recommender systems depend on. Beyond regulatory requirements, Māori data governance principles emphasise that data is taonga, requiring culturally aligned protection, collective benefit, and contextual integrity (Hudson, M. and Beaton, A. and Milne, M. and Port, W. and Russell, K. and Smith, B. and Toki, V. and Uerata, L. and Wilcox, P., 2016; Raraunga, 2018). However, media organisations often use behavioral data and third-party analytics without such safeguards, creating ethical and cultural risks. Although MBIE Responsible AI Guidance for Businesses promotes governance structures and accountability (MBIE, 2025a), its voluntary nature leads to inconsistent adoption. Evidence suggests that governance and audit practices lag behind AI implementation (A. I. Forum, 2025). This creates a structural accountability gap where opaque data practices reduce user trust, as individuals become uncertain about how their data is collected, used, and governed (Wong et al., 2023).

The transfer of these principles to media personalisation requires justification rather than assumption. Te Mata Ira (Hudson, M. and Beaton, A. and Milne, M. and Port, W. and Russell, K. and Smith, B. and Toki, V. and Uerata, L. and Wilcox, P., 2016) was developed for genomic research with Māori, a context of explicit consent, identifiable whakapapa and high cultural sensitivity. Behavioural data from news consumption differs in each respect: it is collected passively, is rarely framed as research, and is not ordinarily labelled by ethnicity. It does not follow that all behavioural data held by a media organisation constitutes Māori data. What follows is narrower. Where a recommender system infers ethnicity, cultural affiliation or interest in Māori content, the resulting data relates to Māori and engages Māori data interests. The principles of kotahitanga and kaitiakitanga articulated by Raraunga (2018) also reach questions the Privacy Act does not, including whether data use produces collective benefit and who exercises guardianship over it. Māori data sovereignty, Māori data governance, privacy law and Te Tiriti-informed organisational practice are distinct and are treated separately here.

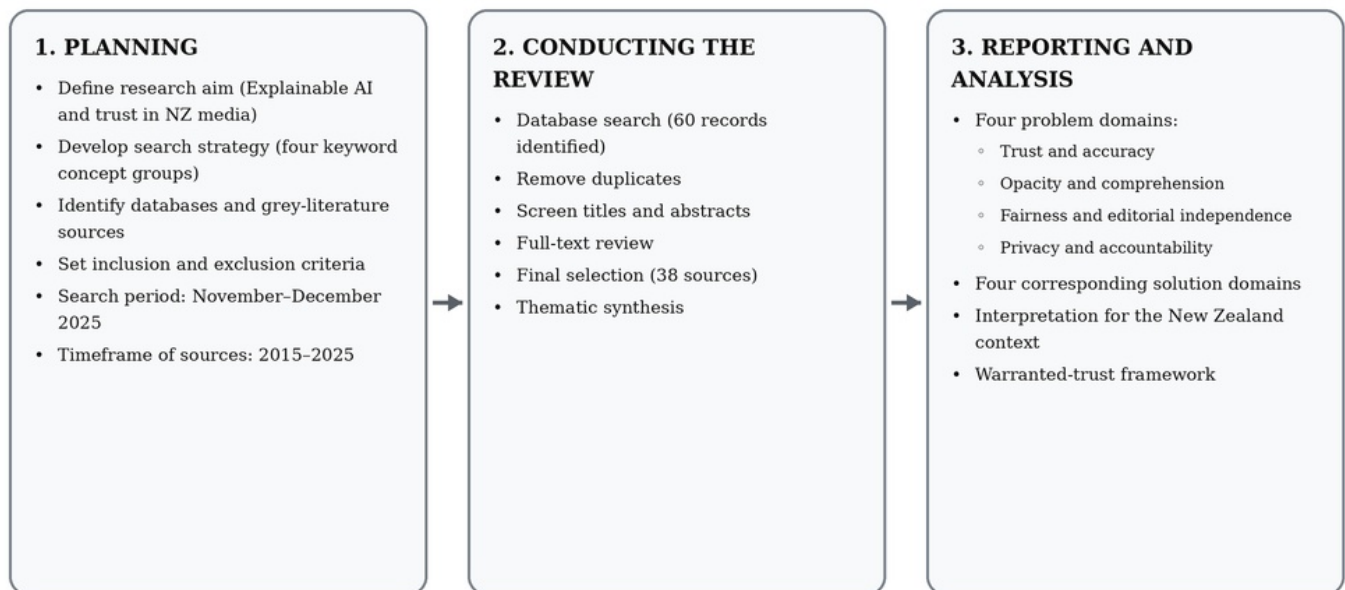
Despite the clear identification of these interconnected challenges, a critical gap remains in defining how these is-

sues can be addressed in an integrated and context-specific manner within the New Zealand digital media landscape. Existing research has largely examined transparency, fairness, or data governance in isolation, with limited attention to how these dimensions can be operationalised collectively to rebuild public trust in algorithmically curated news environments. Furthermore, there are no practical frameworks that align technical explainability with user perceptions, editorial practices, and culturally grounded data governance principles such as Māori data sovereignty. This gap is particularly significant in New Zealand, where trust in the media is already fragile, and societal expectations around ethical AI are still evolving. Therefore, this study aims to examine how explainable AI, combined with robust data ethics and governance mechanisms, can be implemented by New Zealand digital media organisations attempting to rebuild and sustain public trust in AI-driven news recommendation systems.

METHODOLOGY

The present study has chosen a qualitative, desk-based research design, and presents a systematic literature review to examine how XAI may support warranted trust in the digital media sector in New Zealand. The strategy of a literature review was chosen in order to make a synthesis of available academic, policy, and industry literature on AI-based recommender systems, and user trust, systematic and transparent. Figure 1 aimed at addressing conceptual, ethical, and governance concerns, instead of testing the systems empirically, so secondary qualitative research was considered to be the most appropriate.

Figure 1: Methodology for Examining Explainable Artificial Intelligence and Trust in New Zealand Digital Media



Published peer-reviewed academic sources were located in dedicated academic databases (Scopus, Web of Science, IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, Emerald Insight and Google Scholar), each of which was searched separately, together with the arXiv repository for preprint material. Such repositories have been chosen because they are fully covered in terms of explainable AI, recommender systems, fairness-conscious algorithms, and human-computer interaction. Authoritative grey literature was also included to contextualise the inter-

national results in the New Zealand context. This included reports and advice by government agencies, regulators, and research organisations such as MBIE, the Department of Internal Affairs, the Office of the Privacy Commissioner, the Broadcasting Standards Authority, InternetNZ and Stats New Zealand.

The strategy of the search was organised, with the use of combination keywords across four concept groups: explainability terms (explainable AI, XAI, algorithmic transparency, explanation interfaces); system terms (recommender system, news recommendation, personalisation, algorithmic curation); outcome terms (trust, media trust, user perceptions, fairness, viewpoint diversity); and governance terms (data ethics, privacy, accountability, digital governance, Māori data sovereignty). Terms were combined with OR within groups and AND across groups, with New Zealand added as a locational qualifier for grey-literature searches. Searching was conducted during November and December 2025. The process of source selection included abstract and full-text selections in terms of their relevance, credibility, and suitability for the research objectives. Priority was given to those articles published since 2015 because the researchers needed to include the latest events in AI technologies, media disruption, and responses driven by regulation.

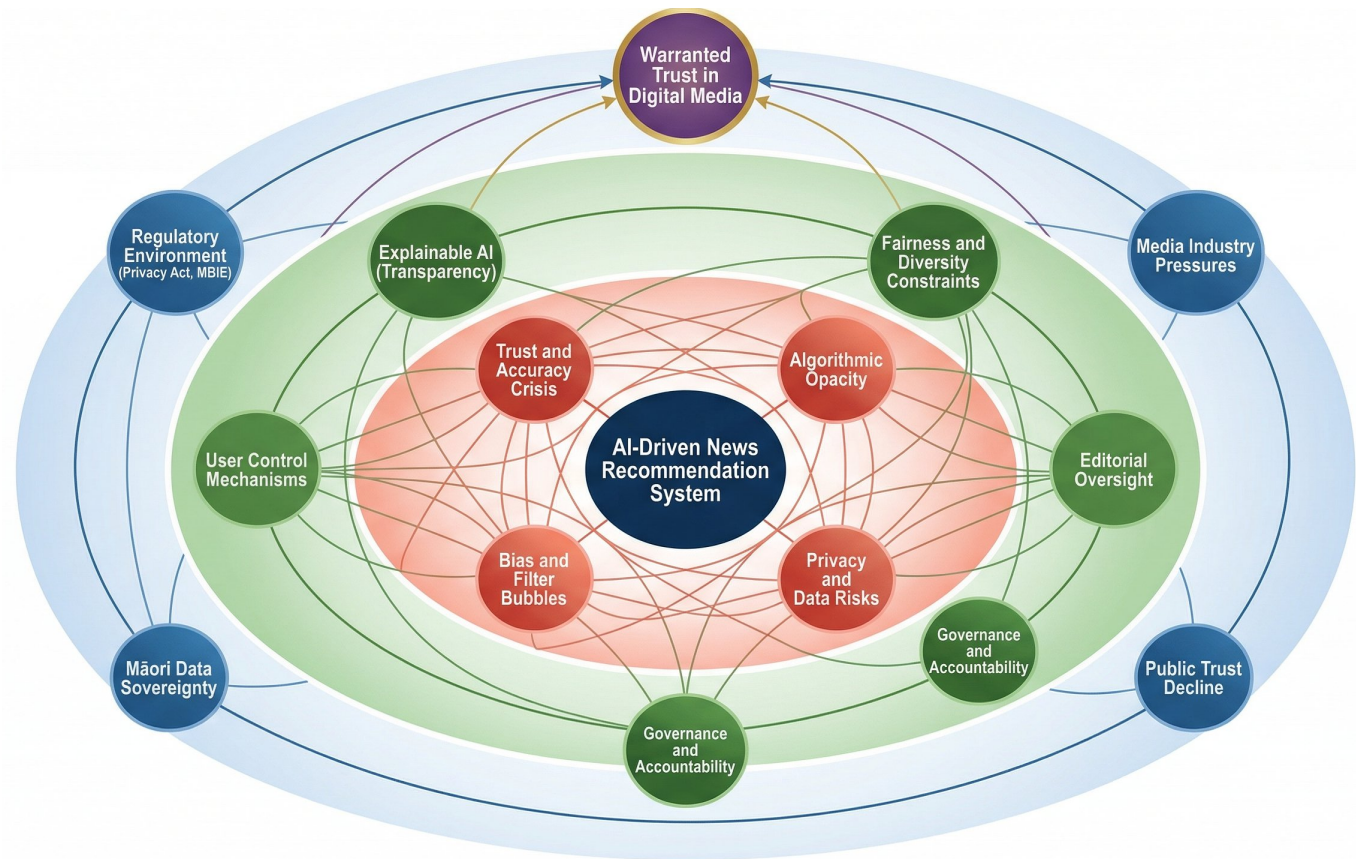
The study employed thematic synthesis of the shortlisted literature to identify recurring patterns related to transparency, trust, accountability, and culturally grounded governance. Synthesis was conducted by the first author, with the coding framework and thematic structure reviewed and agreed by the co-authors; formal double-coding and inter-coder reliability testing were not undertaken. Coding proceeded inductively: descriptive codes were assigned to findings and claims within each source, related codes were grouped into candidate categories, and these were consolidated into the four problem domains and four corresponding solution domains reported in the Literature Review and Findings and Discussion sections. Where sources disagreed, most notably on whether explanations reliably increase trust, the divergence is reported in the findings rather than resolved, and the more cautious interpretation is adopted. The initial database search identified 60 sources, which were subsequently refined through abstract screening and full-text review based on relevance, credibility, publication date, and alignment with the study's focus on explainable AI, recommender systems, and public trust in digital media. This screening process resulted in a final set of 38 academic and policy-oriented sources that formed the basis of the systematic literature review. Inclusion criteria were publication from 2015 onwards to capture recent developments in AI and digital media; direct relevance to at least one of the four concept groups; publication in English; and, for grey literature, authorship by a government agency, regulator, industry body or established research centre. Exclusion criteria removed material published before 2015 unless foundational to a concept used in the review, sources with no substantive engagement with any concept group, and promotional or vendor material. Counts of duplicates removed and of records excluded at each screening stage were not recorded at the time and are not reconstructed here. The resulting themes were synthesised to support the research objectives and to develop a contextualised understanding of how XAI can strengthen trust in New Zealand's digital media ecosystem 2.

Methodological Limitations and Ethical Statement

Four limitations of the review should be read alongside the findings. Stage-level counts of duplicates removed and records excluded at each screening stage were not recorded at the time, so the screening sequence is reported descriptively rather than as a full record. No formal quality-appraisal instrument was applied; source credibility was assessed through peer-review status, institutional authority and methodological transparency, and three retained sources are preprints that have not completed peer review. Coding was undertaken by the first author with co-author review rather than independent double-coding. The retained sources are predominantly international, and

the review located no empirical New Zealand study of user responses to algorithmic news recommendation, so claims about New Zealand are limited to what New Zealand sources establish. This study used publicly available secondary sources and did not involve human participants or the collection of personal data, so formal human-participant ethics approval was not required, as desk-based research using publicly available secondary sources does not fall within the categories requiring ethics application under the relevant institutional policy. Māori data sovereignty concepts are engaged here as published principles; the authors do not speak on behalf of Māori.

Figure 2: Conceptual Framework of XAI, Digital Media Trust, and Governance Factors in New Zealand



Findings and Discussion

The findings of the systematic literature review reveal four interconnected solution domains that directly respond to the key challenges identified in the previous section. The thematic synthesis highlights that rebuilding trust in AI-driven news environments requires an integrated approach centered on transparency, fairness, privacy, and accountable governance. Rather than isolated interventions, the literature consistently emphasises that these dimensions must be operationalised collectively to support what is defined as warranted trust with trust grounded in evidence, explainability, and institutional responsibility rather than passive acceptance (Wong et al., 2023). The following discussion presents these solution domains as evidence-based responses derived from the reviewed literature, aligned with emerging national expectations for responsible AI use in New Zealand.

Throughout this section, three levels of claim are distinguished: what the reviewed literature found, how the

authors interpret that finding for the New Zealand context, and what organisational action the authors propose. Proposed actions are evidence-informed recommendations that have not been tested in New Zealand news organisations.

Rebuilding Public Trust and Confidence in Accuracy

The systematic literature review identifies strengthening trust in news accuracy as a central requirement in algorithmically mediated media environments. The findings indicate that perceptions of sensationalism, bias, and declining editorial standards can be mitigated through enhanced transparency and accountability mechanisms. The reviewed literature establishes that perceptions of weakening editorial standards are associated with lower trust in news. We interpret this as indicating that demonstrated accountability matters more than asserted trustworthiness, and propose on that basis that public-facing trust indicators, such as dashboards displaying correction rates, upheld regulatory decisions and editorial clarifications, could strengthen perceived institutional accountability. No reviewed study evaluates such a mechanism (Authority, 2024; Myllylahti & Treadwell, 2025).

The literature further demonstrates that incorporating quality-weighting mechanisms within recommendation systems enables prioritisation of verified and high-trust content over engagement-driven outputs, aligning algorithmic processes with journalistic values (Diakopoulos & Koliska, 2017; New Zealand On Air, 2024). We further propose that transparent editorial integrity practices, including disclosure of AI-assisted content and structured correction processes, could support perceptions of procedural fairness. This is an author recommendation rather than a finding of the reviewed studies. Collectively, these findings suggest that reinforcing accuracy and accountability is fundamental to restoring trust in algorithmic news systems.

Enhancing Transparency and Improving User Comprehension

The literature consistently highlights transparency and user comprehension as critical determinants of trust in AI-driven recommendation systems. Findings from the review indicate that the absence of clear explanations regarding content selection contributes significantly to user scepticism and perceived opacity. Survey research across five countries, none of them New Zealand, finds that users want both transparency and control in news recommender systems (Mitova et al., 2024), and that perceptions of these systems relate to trust in the outlets deploying them (Blassnig et al., 2024). Miller (2019) argues on conceptual grounds that explanations aid understanding only where they align with human reasoning.

Explainable AI research further demonstrates that contrastive, narrative, and example-based explanations are more effective than technical descriptions, as they align with human reasoning processes and enhance comprehension (Kazmierczak et al., 2026; Liao & Varshney, 2021; Tintarev & Masthoff, 2015). The findings also indicate that user control features, including the ability to adjust preferences or reset personalisation, are associated with greater perceived agency and trust in a United Kingdom study (Wong et al., 2023), with comparable findings in adjacent e-commerce contexts (Acharya et al., 2023). Overall, the literature suggests that combining transparency with user control mechanisms is essential to both reducing perceived manipulation and strengthening trust in personalised news environments.

Safeguarding Fairness, Viewpoint Diversity and Editorial Independence

The systematic review identifies fairness and viewpoint diversity as critical challenges in algorithmic news recommendation systems. The findings indicate that, without intervention, recommender systems tend to overexpose popular

content while marginalising minority perspectives, thereby reinforcing existing inequalities (Biega et al., 2018; Jin et al., 2023). This effect is particularly significant in the New Zealand context, where concerns about underrepresentation of Māori and Pasifika communities and perceptions of media bias are already prevalent (Hope, W., Baker, S., Daniels, R., Hoar, P., McEwan, R., Myllylahti, M., Mountfort, P., Middleton, A., Selmán Julian, D., Treadwell, G. & Watts, T., 2025; Kantar, 2021).

The literature demonstrates that integrating fairness-aware constraints and diversity weighting mechanisms can help ensure balanced content exposure and mitigate filter bubble effects (Areeb et al., 2023). Additionally, maintaining a visible editorial layer, such as curated sections or public interest content, reinforces the role of human judgement and supports editorial independence (Diakopoulos & Koliska, 2017). Transparency in editorial AI policies and periodic diversity audits are also identified as mechanisms to enhance perceived fairness. These findings suggest that fairness and diversity are essential to sustaining trust in algorithmic news systems.

Strengthening Privacy, Data Ethics and Accountability

The literature identifies privacy, data ethics, and accountability as foundational to trust in AI-driven personalisation systems. Findings indicate that high levels of public concern regarding data security and misuse create significant barriers to trust (InternetNZ, 2024). Regulatory frameworks such as the Privacy Act 2020 establish requirements for transparency, data minimisation, and accountability in data handling (Commissioner, 2023).

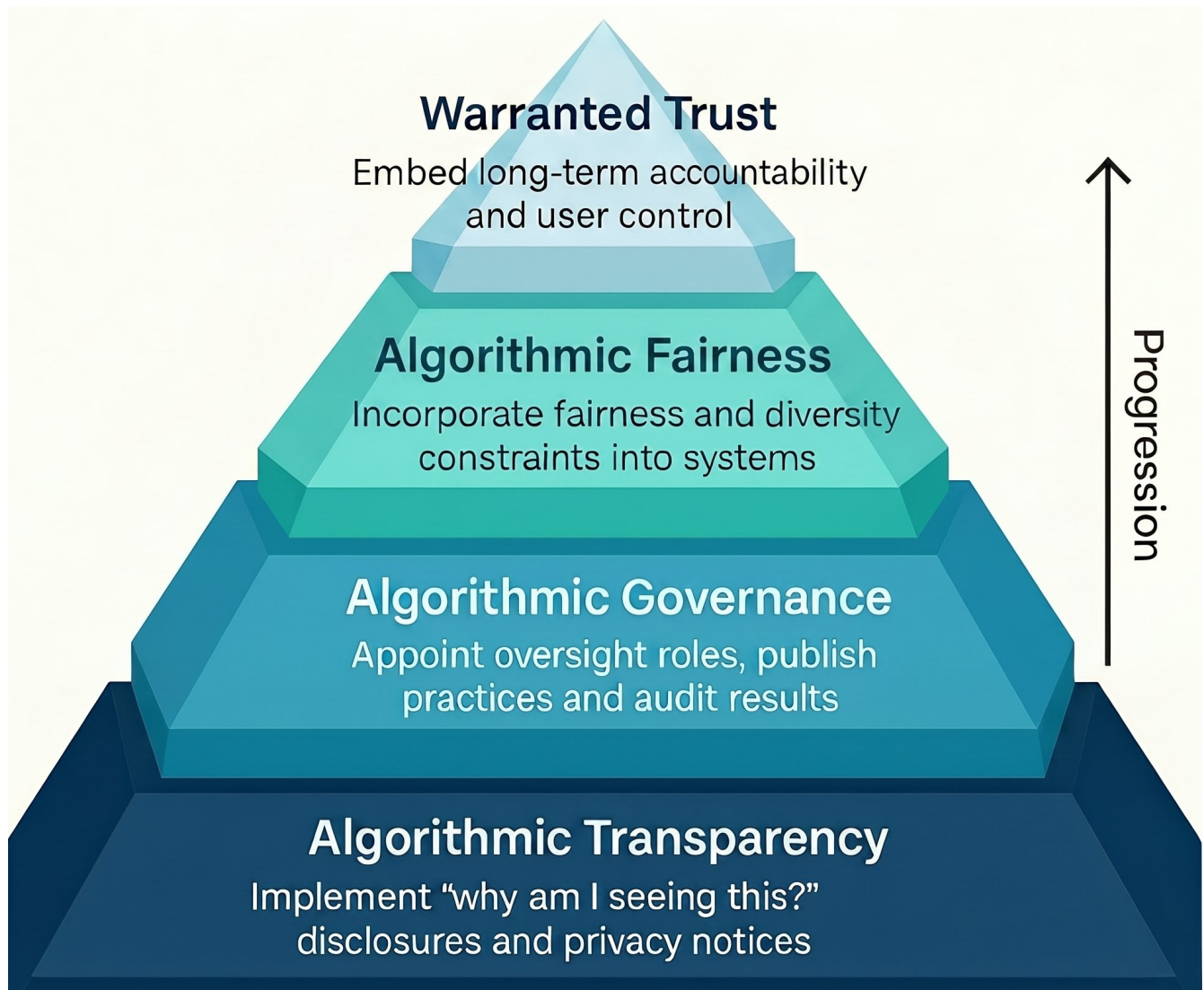
Beyond regulatory compliance, the literature emphasises the importance of culturally grounded data governance, particularly Māori data sovereignty principles, which require data to be treated as taonga and managed in ways that ensure collective benefit and contextual integrity (Hudson, M. and Beaton, A. and Milne, M. and Port, W. and Russell, K. and Smith, B. and Toki, V. and Uerata, L. and Wilcox, P., 2016; Raraunga, 2018). The review also highlights gaps in implementation, as governance frameworks and auditing practices often lag behind AI adoption (A. I. Forum, 2025). Research shows that lack of visibility into data practices reduces user trust (Ge et al., 2024; Wong et al., 2023). Overall, the findings indicate that embedding transparent governance and ethical data practices is essential to closing accountability gaps and rebuilding trust.

Recommendations and Implications

The solutions outlined in the previous section can be translated into a strategic set of recommendations intended to help New Zealand digital media organisations restore public confidence in AI driven news recommendation systems. As shown in Figure 3, trustworthy AI development follows a progressive pathway beginning with basic transparency measures and extending through governance, fairness and ultimately long term warranted trust. This progression is essential in New Zealand where declining trust, high digital media consumption and significant concerns about privacy and fairness create complex environments for deployment of recommendation technologies. The recommendations presented here therefore aim to strengthen public confidence through practical, evidence based and culturally aligned measures that align with national expectations for responsible AI use.

News organisations in the short term can provide immediate benefits in terms of the public trust with improved transparency and understanding. One of most effective measures is introduction of 'Why am I seeing this' explanations for recommended articles which provide users with simple, contextual insights into why specific stories appear in their feed. Research shows that even short, well-timed explanations can motivate less suspicion and better perceptions of fairness by helping users to understand that decisions of AI algorithms reflect signals that they can identify

Figure 3: Roadmap for Building Trustworthy Recommendation Systems



rather than hidden manipulation (Blassnig et al., 2024; Miller, 2019). These explanations should be accompanied by plain language data notices that clarify which types of behavioral information are used, how long they are retained for, and how users can limit or opt out of personalised tracking. This is particularly important in New Zealand, where 65% of the population reports high concern about the security of personal data (InternetNZ, 2024). Providing users with basic personalisation controls such as ability to adjust topic preferences or reset their recommendation history further supports perceptions of fairness and autonomy. Simultaneously, visible editorial integrity messaging including explanations of how corrections are handled or how AI assisted content is reviewed, gives audiences confidence that journalistic values remain central to organisational identity. Collectively, these actions build a basis for trust by showing openness and respect for user agency.

Organisations can build trust in the medium term by establishing formal governance mechanisms that include fairness protections and editorial safeguards around algorithmic systems. Practices such as establishing internal gover-

nance frameworks including model registers with designated oversight roles and documentation of data sources and risk controls align organisational practice with emerging expectations set out in Algorithm Charter for New Zealand and MBIE Responsible AI Guidance (Department of Internal Affairs, 2020; MBIE, 2025). Such structures offer comfort in knowing that recommendation systems are treated with the same care as applications that are public sector and rely on AI. These governance commitments should be supplemented by periodic fairness and diversity audits of representation of, for example, Māori, Pasifika, regional content and other groups across recommended content. This step is important in light of concerns about under representation identified in audience studies (Kantar, 2021). Maintaining a visible human editorial layer such as manually curated editor selections within personalised environments further reinforces that journalistic judgment remains active by countering perceptions that algorithms have displaced editorial independence. Publication of clear accessible policies outlining where AI is employed in content creation or ranking also helps to increased trust by reducing uncertainty between human and algorithmic decision making. Long term recommendations focus on embedding trust as a sustained organisational value rather than a temporary initiative. Publishing annual transparency and fairness reports allows media organisations to communicate how their systems are performing with respect to accuracy, representation, privacy safeguards and risk mitigation. Regular external or independent reviews of algorithmic systems, governance processes and privacy protections add a further layer of credibility, particularly when such reviews are publicly released. Sector wide collaboration across NZME, Stuff, RNZ, TVNZ and relevant industry bodies could also establish shared norms and minimum standards for trustworthy AI in the New Zealand media environment. Such alignment reduces inconsistency across organisations and strengthens media sector collective legitimacy. Finally, long term trust requires investing in public literacy and community engagement initiatives that help audiences understand how algorithms work, what privacy protections exist and how they can make informed decisions about their news consumption. These efforts can reduce misinformation while enhancing civic understanding and foster greater resilience in the face of evolving AI technologies.

Overall, recommendations presented here reinforce the staged development illustrated in Figure 3. New Zealand media organisations can demonstrate credible commitment to responsible AI use by embedding transparency, user comprehension, fairness, privacy protection, Māori data ethics and structured governance into every stage of algorithmic design and deployment. Such practices support the emergence of warranted trust which is a form of public confidence grounded in evidence, accountability and respect for user rights, rather than reliance on opaque technological processes (de Brito Duarte et al., 2023). This approach is essential for safeguarding the democratic role of news media in New Zealand as AI driven personalisation continues to expand.

Conclusion

The increasing integration of AI driven news recommendation systems within the New Zealand digital media sector is occurring at a time of declining public trust with heightened concern about privacy and fairness and growing uncertainty about how algorithms influence civic information. This report has shown that trust cannot be rebuilt through technology alone, but it requires transparency and accountable governance as well as culturally grounded data practices and clear evidence that personalisation supports rather than undermines journalistic integrity. By embedding long term commitments to accuracy, fairness, user control and responsible AI governance, New Zealand media organisations can foster warranted trust and strengthen their democratic role in New Zealand's evolving digital information environment.

REFERENCES

- A. I. Forum, N. Z. (2025). *AI in action: Productivity and performance in New Zealand businesses*. https://aiforum.org.nz/wp-content/uploads/2025/03/AI-in-Action_March2025-Report-compressed.pdf
- Acharya, N., Sassenberg, A.-M., & Soar, J. (2023). Consumers' behavioural intentions to reuse recommender systems: Assessing the effects of trust propensity, trusting beliefs and perceived usefulness. *Journal of Theoretical and Applied Electronic Commerce Research*, 18(1), 55–78. <https://doi.org/10.3390/jtaer18010004>
- Alam, S., Rehman, S., Pawade, A., & Kularatne, I. (2026). *Understanding the "Why": A Cross-Platform Evaluation of Explainable AI Features in Social Media*. <https://aisel.aisnet.org/confirm2026/3>
- Areeb, Q. M., Nadeem, M., Sohail, S. S., Imam, R., Doctor, F., Himeur, Y., Hussain, A., & Amira, A. (2023). Filter bubbles in recommender systems: Fact or fallacy—A systematic review [Article e1512]. *WIREs Data Mining and Knowledge Discovery*, 13(6). <https://doi.org/10.1002/widm.1512>
- Authority, B. S. (2024). *Annual report 2024*. <https://www.bsa.govt.nz/assets/BSA-Annual-Report-2024.pdf>
- Biega, A. J., Gummadi, K. P., & Weikum, G. (2018). Equity of attention: Amortizing individual fairness in rankings. *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, 405–414. <https://doi.org/10.1145/3209978.3210063>
- Blassnig, S., Mitova, E., Strikovic, E., Urman, A., de Vreese, C., Hannák, A., & Esser, F. (2024). User perceptions of news recommender systems and trust in media outlets: A five-country study. *Journalism Studies*, 25(10), 1182–1204. <https://doi.org/10.1080/1461670X.2024.2364628>
- Carney. (2025). *Digital 2025: New Zealand*. <https://indd.adobe.com/view/9e96312a-78fo-415e-8dff-a07c93fc9c14>
- Commissioner, P. (2023). *Artificial intelligence and the information privacy principles*. <https://www.privacy.org.nz/assets/New-order/Resources-/Publications/Guidance-resources/AI-Guidance-Resources-/AI-and-the-Information-Privacy-Principles.pdf>
- Data.Govt.NZ. (2020). *Algorithm charter for Aotearoa New Zealand*. <https://data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter>
- DataReportal. (2025). *Digital 2025: New Zealand*. <https://indd.adobe.com/view/9e96312a-78fo-415e-8dff-a07c93fc9c14>
- de Brito Duarte, R., Correia, F., Arriaga, P., & Paiva, A. (2023). AI trust: Can explainable AI enhance warranted trust? [Article 4637678]. *Human Behavior and Emerging Technologies*, 2023. <https://doi.org/10.1155/2023/4637678>
- Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, 5(7), 809–828. <https://doi.org/10.1080/21670811.2016.1208053>
- FigureNZ. (2023). *Concerns about artificial intelligence (AI) among people in New Zealand*. <https://figure.nz/chart/pbEckovsM6VscVop>
- Ge, Y., Liu, S., Fu, Z., Tan, J., Li, Z., Xu, S., Li, Y., Xian, Y., & Zhang, Y. (2024). A survey on trustworthy recommender systems [Article 13]. *ACM Transactions on Recommender Systems*, 3(2). <https://doi.org/10.1145/3652891>
- Hope, W., Baker, S., Daniels, R., Hoar, P., McEwan, R., Myllylahti, M., Mountfort, P., Middleton, A., Selman Julian, D., Treadwell, G. & Watts, T. (2025). *AOTEAROA New Zealand Media Ownership 2025*. <https://www.jmadresearch.com/new-zealand-media-ownership>
- Hudson, M. and Beaton, A. and Milne, M. and Port, W. and Russell, K. and Smith, B. and Toki, V. and Uerata, L. and Wilcox, P. (2016). *Te Mata Ira: Guidelines for genomic research with Māori: Guidelines for genomic research with Māori*. University of Waikato. <https://www.waikato.ac.nz/assets/Uploads/Research/Research-institutes-centres-and-groups/Centres/MIGC/Te-Mata-Ira-Genome-Research-Guidelines.pdf>
- InternetNZ. (2024). *New Zealand's internet insights 2024*. Prepared by Verian. <https://internetnz.nz/assets/Archives/New-Zealands-Internet-Insights-2024.pdf>
- Jin, D., Wang, L., Zhang, H., Zheng, Y., Ding, W., Xia, F., & Pan, S. (2023). A survey on fairness-aware recommender systems [Article 101906]. *Information Fusion*, 100. <https://doi.org/10.1016/j.inffus.2023.101906>
- Kantar. (2021). *Strong public media audience data analysis*. Prepared for the Ministry for Culture; Heritage. <https://www.mch.govt.nz/sites/default/files/2023-11/kantar-report-strong-public-media-audience-0722.pdf>

- Kazmierczak, R., Azzolin, S., Berthier, E., Hedström, A., Delhomme, P., Filliat, D., Bousquet, N., Frehse, G., Mancini, M., Caramiaux, B., Passerini, A., & Franchi, G. (2026). Benchmarking XAI explanations with human-aligned evaluations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(44), 37491–37500. <https://doi.org/10.1609/aaai.v40i44.41082>
- Liao, Q. V., & Varshney, K. R. (2021). *Human-centered explainable AI (XAI): From algorithms to user experiences [Preprint]*. *arXiv*. <https://doi.org/10.48550/arXiv.2110.10790>
- MBIE. (2025a). *New Zealand strategy for artificial intelligence: Investing with confidence*. <https://www.mbie.govt.nz/business-and-employment/economic-growth/digital-policy/new-zealands-ai-strategy-investing-with-confidence>
- MBIE. (2025b). *Responsible AI guidance for New Zealand organisations*. <https://www.mbie.govt.nz/assets/responsible-ai-guidance-for-businesses.pdf>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Ministry for Culture and Heritage. (2024). *Media and broadcasting sector profile 2024*. <https://www.mch.govt.nz/sites/default/files/2025-03/media-industry-profile-report-2024.pdf>
- Mitova, E., Blassnig, S., Strikovic, E., Urman, A., de Vreese, C., & Esser, F. (2024). Exploring users' desire for transparency and control in news recommender systems: A five-nation study. *Journalism*, 25(10), 2001–2021. <https://doi.org/10.1177/14648849231222099>
- Myllylahti, M., & Treadwell, G. (2025). *Trust in news in Aotearoa New Zealand 2025* [AUT Research Centre for Journalism, Media and Democracy]. <https://openrepository.aut.ac.nz/items/d2012dda-93f1-4e4b-afda-b549fad8b1f6>
- New Zealand On Air. (2024). *Where are the audiences?* <https://www.nzonair.govt.nz/news/where-are-the-audiences-2024/>
- NZME. (2025). *NZME Limited 2024 full year financial results*. <https://www.nzme.co.nz/news/nzme-limited-2024-full-year-financial-results>
- Raraunga, T. M. (2018). *Principles of Māori data sovereignty (Brief No. 1)*. <https://www.temanararaunga.maori.nz/tutohinga>
- Stats New Zealand. (2024). *New Zealanders' trust in key institutions declines*. <https://www.stats.govt.nz/news/new-zealanders-trust-in-key-institutions-declines/>
- Strategic framework: Better digital future for Aotearoa*. (2022). <https://www.mbie.govt.nz/business-and-employment/economic-growth/digital-policy/digital-technologies-industry-transformation-plan-2/strategic-framework>
- Sunny, A. D. (2025). *Preliminary quantitative study on explainability and trust in AI systems [Preprint]*. *arXiv*. <https://arxiv.org/abs/2510.15769>
- Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (2nd ed., pp. 353–382). Springer. https://doi.org/10.1007/978-1-4899-7637-6_10
- Wong, Y. N., Jones, R., Das, R., & Jackson, P. (2023). Conditional trust: Citizens' council on data-driven media personalisation and public expectations of transparency and accountability. *Big Data & Society*, 10(2). <https://doi.org/10.1177/20539517231184892>
- Zealand, R. N. (2024, March 5). *Cuts and closures in New Zealand news media industry: What you need to know*. Retrieved March 5, 2024, from <https://www.rnz.co.nz/news/what-you-need-to-know/511466/cuts-and-closures-in-new-zealand-s-news-media-industry-what-you-need-to-know>