Understanding and Countering Dis/Misinformation from a Multilingual and Minority Language Perspective

Iker Erdocia and Derek Walsh, Dublin City University

Funded by Research Ireland New Foundations

2026

DCU Institute of Future Media, Democracy and Society

A report by Iker Erdocia and Derek Walsh

July 2026

About the authors

Iker Erdocia, PhD, is Assistant Professor and Director of Research in the School of Applied Language and Intercultural Studies (SALIS) in Dublin City University. He is the Principal Investigator (PI) of the MULTIDISIN project. He is also the PI for the Horizon Europe FOSTERLANG project on minority languages in Europe and the AHRC TransCoLL project on cultures of language learning in the UK and Ireland. His research sits at the interface of language, policy and politics. Iker has published widely in high-ranking journals within the fields of sociolinguistics and language policy.

Derek Walsh is a writer and researcher based in Dublin. He holds an MA in Social Media Communications from Dublin City University.

Acknowledgements

We gratefully acknowledge the support of Dublin City University (DCU), the Institute for Future Media, Democracy and Society (FuJo) and the ADAPT Centre.

We also wish to thank Brendan Spillane (Co-Principal Investigator of MULTIDISIN, University College Dublin); the presenters in the MULTIDISIN Webinar Series (Silje Alvestad, Rodrigo Cetina, Susan Daly, Asif Ekbal, Rina Kumari and Carolina Scarton) and the workshop presenters (Rodrigo Cetina, Encarnación Hidalgo Tenorio, Jakub Kubś, Aidan O'Brien and Craig Willis) for sharing their expertise and insights, which have informed this report. We also thank all webinar and workshop participants for their valuable contributions and engagement.

Finally, we acknowledge with gratitude the financial support of Research Ireland, whose funding made this project possible.

Table of Contents

Executive summary

This report is an output of the project Understanding and Countering Dis/Misinformation from a Multilingual and Minority Language Perspective (MULTIDISIN), funded by Research Ireland New Foundations (Strand 1c, Interdisciplinary Research Networking Awards).

While generative AI offers opportunities to enhance democratic processes, it also presents numerous risks to democracies. MULTIDISIN project seeks to mitigate the harmful societal impacts of the digital transition by understanding and countering dis/misinformation from a comparative European perspective, with particular attention to medium-sized and minority languages, especially Irish. By bringing together (socio)linguists, minority language scholars, computer scientists, journalists, fact-checkers and policy specialists, the project has opened a space of dialogue and established a network of researchers contributing to the development of conceptual and AI tools to track disinformation narratives in Irish and other under-researched European languages.

The report draws on a review of the research literature in cross-lingual dis/misinformation detection; on the project's five-part webinar series, which brought perspectives from natural language processing (NLP), linguistics, journalism and minority-language scholarship; and on an expert round-table workshop convened by the project with participants from disinformation monitoring, digital regulation, social sciences, NLP and minority-language media research. Section 1 sets out the interdisciplinary background. Section 2 explains the societal and political importance of dis/misinformation viewed from multilingual and minority-language perspectives. Section 3 reviews the state of the art across five areas (linguistics, NLP, journalism and media, political science and interdisciplinary research), integrating the findings of the project webinars. Section 4 draws out implications for academics and policymakers, and Section 5 proposes a future research agenda. Recommendations are grouped below by target audience and repeated in the relevant sections.

Summary and recommendations

Dis/misinformation is not experienced equally across languages. Of the world's 7,000+ living languages, roughly 90% are considered low-resource; only around one hundred have meaningful representation in the datasets used to train AI systems, and a large majority of online content is produced in a handful of dominant languages, particularly English. The consequence, documented across every strand of evidence reviewed for this project, is a structural protection gap: the languages least served by platform content moderation, fact-checking infrastructure and language technology are also the languages in which false and manipulative content can circulate longest without detection or correction. Speakers of minority, minoritised and medium-sized languages are therefore doubly exposed: first as targets of identity-based disinformation, and second as communities that the traditional media and counter-disinformation ecosystem are least equipped to defend.

As was often highlighted during the project activities, generative AI sharpens both edges of this problem. Large language models (LLMs) lower the cost of producing fluent, localised and personalised disinformation in many languages at once, while performing worst in output quality, safety-filter robustness and detection accuracy, precisely in low-resource languages. At the same

time, these models offer genuinely new capabilities for defenders, enabling cross-lingual transfer of detection tools to languages for which no dedicated resources exist. Whether LLMs prove friend or foe to minority-language communities depends, in large part, on political and policy choices.

Ireland is a revealing case. As an Anglophone country, it imports mis/disinformation from the much larger US and UK information spheres without any translation lag, while Irish (Gaeilge) and significant community languages, such as Polish, sit largely outside the State's counter-disinformation, media-literacy and public-information efforts. Ireland currently lacks a permanent, systematic research infrastructure for monitoring the volume, sources and targets of disinformation across languages. As a result, policymakers lack a robust domestic evidence base on the scale and characteristics of the problem. In order to address this gap, the recommendations below address the European, Irish, research and platform levels.

For EU policymakers and regulators

R1. A more granular and systematic language-disaggregated transparency should be required under the Digital Services Act (DSA): systemic risk assessments and mitigation reporting by very large online platforms (Articles 34–35) should explicitly identify minority-language and identity-based harms, and transparency reports should disaggregate content-moderation performance, resourcing and outcomes by language.

R2. Platform moderation resources should be allocated in proportion to the risk of harm to a linguistic community, not to its market size, and regulators should treat persistent failure to moderate non-official languages as a systemic risk under the DSA.

R3. The monitoring of foreign information manipulation and interference (FIMI) should treat language as a targeting vector in its own right, progressively extending the identity-based approach adopted by the European External Action Service to minority languages.

R4. Implementation of the AI Act should include multilingual safety evaluation: general-purpose model providers should be required to evaluate and disclose model performance and safety-filter robustness across EU languages, and certification pathways should be funded for AI tools in low-resource and minority languages.

For Irish policymakers

R5. Ireland should fund a permanent, independent research and monitoring capability on dis/misinformation, with an explicit mandate to establish baseline evidence on the Irish-language information environment, building on and extending the National Counter Disinformation Strategy.

R6. Public information campaigns and media-literacy programmes should reach beyond English and Irish: mandatory provision in Irish should be complemented by proportionate provision in widely spoken community languages, delivered in partnership with trusted community intermediaries.

R7. The State should invest in Irish-language data resources, such as corpora, annotated datasets and an open benchmark for dis/misinformation detection, as public digital infrastructure, so that Irish is not dependent on commercial priorities for its presence in language technology.

For researchers, funders and higher education institutions

R8. Funders should support the construction of multilingual benchmarks and community-governed datasets for low-resource languages, and researchers should systematically document and publish where LLMs and detection systems fail in these languages.

R9. Research on dis/misinformation in minority languages should be interdisciplinary by design, pairing computational approaches with sociolinguistic, ethnographic and journalistic expertise, and building durable partnerships among researchers, fact-checkers and communities.

R10. Higher education institutions in the EU and beyond have an obligation to sustain independent research on generative AI safety and multilingual disinformation, particularly where commercial actors will not; funding instruments should extend beyond short project cycles, such as the present one, to preserve capacity.

For platforms and AI developers

R11. Platforms should publish language-specific moderation data and employ more moderators with expertise in non-official languages and the cultural competence needed to serve the communities that use them. They should also improve transparency on their use of machine translations or LLM models, adapt their disinformation classification and crisis-response frameworks to low-resource languages, and make researcher data access under DSA Article 40 genuinely usable for minority-language research. AI developers should extend safety training and evaluation to low-resource languages and co-design deployments with the affected language communities.

1. Background

Misinformation refers to false or misleading content shared without intent to harm; in contrast, disinformation is false or misleading content created, presented and disseminated deliberately, whether for political, financial or social ends. Both circulate through news media, social platforms and, increasingly, closed messaging applications, in forms ranging from fabricated news and manipulated images to rumour, clickbait, propaganda and decontextualised genuine content. The distinction matters for policy (intent shapes both the appropriate response and the legal basis for intervention), but in practice the two phenomena travel together and are addressed here jointly as dis/misinformation (Culloty and Suiter, 2021).

The study of dis/misinformation is inherently interdisciplinary. Linguistics provides the basis for an understanding of how deceptive and hateful language, texts and discourses are constructed — their registers, metaphors, implicatures and audience design. NLP contributes to the detection and generation of technologies that now mediate most of the information environment. Journalism and media studies contribute to the practice of verification and the economics of trust. Political science focuses on the analysis of regulation, election integrity and platform governance. No single discipline can explain, let alone counter, a phenomenon that is simultaneously textual, technical, commercial and political — a conviction that shaped the composition of this project's webinar series and workshop.

Two developments define the current moment. The first is the arrival of generative AI. LLMs can draft, translate, localise and personalise content at negligible cost, and multimodal systems extend this to images, audio and video. Scholarly opinion is divided on how much this changes the underlying dynamics of misinformation: some researchers argue that fears are overblown because the supply of misinformation has never been the binding constraint on its influence (Simon, Altay and Mercier, 2023), while others document concrete new capabilities for manipulation at scale (Chen and Shu, 2024; Bontcheva et al., 2024). What is not in dispute is that the technology shifts the balance of advantage between languages: it strengthens some actors and capabilities while weakening others, and it does so unevenly across the world's languages, as Section 3 details. The second development is regulatory: the EU's Digital Services Act (DSA) and AI Act have created, for the first time, binding obligations on very large platforms and AI providers, obligations whose value for minority-language communities depends entirely on how they are implemented and enforced.

Within this landscape, MULTIDISIN focuses on languages that the research literature in social sciences variously calls minority, minoritised, medium-sized, non-dominant or, in the terminology now standard in NLP, low-resource and under-resourced: languages for which digital text, annotated data, tools and commercial attention are scarce relative to dominant languages. The shift in terminology is itself instructive: it locates the problem not in the size of a speaker community but in the allocation of resources to it. Irish is a paradigm case; it is constitutionally the first official language of Ireland, with statutory support and extensive bilingualism but, nonetheless, occupies a marginal position in the datasets, moderation systems and detection tools that govern the online information environment.

2. Why dis/misinformation matters from multilingual and minority-language perspectives

The content-moderation gap. Platform investment in trust and safety is radically skewed towards English. Figures reported in the project webinar series illustrate the scale: one major video platform reportedly employs over 15,000 English-language content moderators, single figures for some official EU languages, and none for Irish; internal documents disclosed in the Facebook Papers indicated that some 87% of one major platform's budget for classifying misinformation was reportedly devoted to English-language content, though only a small minority of its users are native English speakers. Automated moderation calibrated on dominant-language data may both under-enforce (harmful content circulates unflagged) and over-enforce (legitimate minority-language content is wrongly removed) — a dual risk documented for non-English content analysis generally (Nicholas and Bhatia, 2023). There is at present no direct regulatory requirement on platforms to correct this imbalance.

Identity-based targeting. Language is not only a channel through which disinformation travels; it is a marker of identity around which disinformation is organised. Case evidence gathered by the NATO-supported Barcelona conference on disinformation and minority languages and by the project webinars documents Russian-language disinformation targeting Russian speakers in Ukraine and Moldova; inconsistent moderation between Spanish and Catalan (Aira and Cetina, 2024); digitally isolated communities, such as Santali speakers in India, left without reliable health information during COVID-19; and, in Ireland, disinformation about the Irish language deployed to serve exclusionary definitions of national identity. As reported in the project webinar series, the European External Action Service's FIMI monitoring has, since 2024, drawn attention to disinformation that targets communities based on identity (including language) as a threat category requiring tailored responses.

Democratic stakes. The 2024 Indian general election, discussed in the project webinar on Indian multilingual misinformation, saw AI-generated deepfakes of political leaders circulated across hundreds of WhatsApp groups reaching millions of voters, in multiple languages, with minimal detection resources. In Europe, election-observation work across more than fifteen countries has found that disinformation narratives repeat across borders and languages, while monitoring attention concentrates on each state's dominant language: Mexico's dozens of indigenous languages, and their European counterparts, are simply not watched. Electoral manipulation in an unmonitored language is invisible until its effects are counted.

The Matthew effect in language technology. Market incentives produce more data, better models and more investment for languages that already have them, and exponentially less for those that do not. This systematic inequality has been quantified across the world's languages (Blasi, Anastasopoulos and Neubig, 2022). LLMs inherit and amplify this situation: benchmarks are designed in English, safety training is overwhelmingly English-centric, and performance degrades sharply outside the high-resource core. A literature review presented in the project's opening webinar found that a clear majority of studies assessing LLM performance in under-resourced languages report substantially worse quality, more hallucination on culture-specific names, places and concepts, and weaker disinformation detection; research on safety has shown that translating harmful prompts into

low-resource languages bypasses GPT-4's safeguards in the large majority of attempts (Yong, Menghini and Bach, 2023). Structural social inequality, thus, becomes structural digital inequality, a phenomenon that critics have called a form of digital colonialism.

The Irish double exposure. Ireland experiences both sides of the problem at once. As an Anglophone country, it absorbs US and UK mis/disinformation instantly and without translation friction. Examples include imported anti-abortion advertising in the 2018 referendum, the "great replacement" conspiracy theory localised as a "great plantation", and foreign-operated accounts posing as Irish users during the 2023 Dublin riots. Meanwhile, the Irish language itself occupies an unusual position: because virtually all Irish speakers are bilingual and embedded in the English-language information space, it can be said that there is no isolated Irish-language audience to radicalise, and disinformation appears less in Irish than about Irish, for example, fabricated evidence that Gaeltacht areas are being "overrun" or the language deliberately neglected, manufactured to inflame identity conflict. At the same time, spray-and-pray multilingual operations now reach Irish incidentally: the pro-Russian Pravda network stood up an Irish-language news site as part of a campaign spanning dozens of languages and machine-translated financial scams circulate in visibly non-standard Irish. Importantly, this creates a risk that low-quality synthetic text could enter datasets used to develop LLMs for this language. Community languages such as Polish face a different gap again: statutory public-information spending obligations that protect Irish do not extend to them, leaving large communities dependent on home-country and diaspora media.

Case: the Pravda network's Irish-language outlet

Fact-checkers at The Journal and a member of Ireland's EDMO hub identified an Irish-language news website belonging to the pro-Russian "Pravda" disinformation network, which is part of a coordinated operation replicating content across dozens of languages (see Kubś, 2025). The discovery illustrates two realities of the generative-AI era: automatic translation makes it essentially free to include even very small languages in influence operations, and no monitoring system existed that would have detected an Irish-language operation systematically. The site was found through the informal exchange of a fact-checking network, not through any structured surveillance of the Irish-language information space.

Relevant recommendations: R1–R3, R5, R6.

3. State of the art

This section reviews research on dis/misinformation in multilingual and minority-language contexts across five areas, integrating the contributions of the project webinar series. Each webinar's findings are summarised within the corresponding disciplinary area; the project workshop roundtable informs the political science and interdisciplinary sections.

3.1 Linguistics

Linguistics-oriented research approaches dis/misinformation as an issue of language in use in corpora: how deceptive texts and narratives are constructed, what kind of language and linguistic features they use, how they position audiences, and whether they carry identifiable stylistic signatures. Forensic-linguistic work has demonstrated that fake news can differ measurably from

genuine reporting in linguistic features, register and information density (Grieve and Woodfield, 2023; Põldvere et al. 2023); author-profiling research has shown that the propensity to spread false content can be modelled from language behaviour across languages, with experiments covering English and Spanish (Vogel and Meghana, 2020). Sociolinguistics contributes several decades of research on how identity traits such as age, gender, origin and ideology are indexed in language use, a foundation on which computational profiling of malicious actors is built. Researchers have only recently begun to examine the implications of machine learning and AI for sociolinguistics (e.g. Erdocia, Migge and Schneider, 2024; Erdocia, Schneider and Migge, 2026). In this regard, computational sociolinguists, a distinct subfield within NLP research that pays attention to language use and variation, has emerged in the last three years.

The project webinar on platformed hate speech and fake news presented a decade-long Spanish research programme that operationalises these insights and provides a promising application to minority language contexts (e.g., Francisco et al., 2022). Combining critical discourse analysis, linguistics and AI, the team developed an open-access detection algorithm, which is built on the twin notions of linguistic patterns and "deep relations" between users who interact indirectly. This algorithm identifies radical and disinforming profiles from comparatively small quantities of data, and is now used by Spanish and Catalan police forces and by Europol. Applied to multilingual data (German, English, Spanish, Catalan and Basque) across three case studies — COVID-19, populism and climate change — the research found that disinformation overwhelmingly takes the form of opinion, often laundered through an accompanying news item; that metaphor is the single most frequent deceptive strategy (migration as "tsunami", "invasion"); and that hate speech and disinformation co-occur systematically, targeting migrants above all by legal status, with dehumanising portrayals and deniability strategies such as implicature, ellipsis, scare quotes and humour. The work also produced a taxonomy of deceptive and hateful strategies, subsequently validated against existing classifications.

Two findings from this strand are particularly relevant to minority languages. First, the deceptive strategies identified (metaphor, implicature, emotional appeal) are deeply language- and culture-specific, which is precisely why detection systems trained on English transfer poorly. Second, the linguistic experience of interdisciplinary teams shows that expert annotation pipelines can be extended with non-expert annotators and LLM adjudication, a pragmatic model for languages where expert annotators are scarce.

3.2 Natural language processing

Detection research. Automatic detection of dis/misinformation is a mature research area in English and a fragile one for everything else. Recent surveys of monolingual and multilingual detection for low-resource languages (Wang et al., 2024) converge on the same gaps: scarce annotated datasets, poor cross-lingual transferability, weak coverage of code-mixed language, and vulnerability to AI-generated content. Three families of approach dominate. The first fine-tunes multilingual transformer models across several languages at once; Mul-FaD, covering English, German and French, is a representative example (Ahuja and Kumar, 2023). The second exploits cross-lingual evidence: checking a claim against reporting in other languages measurably improves detection

(Dementieva, Kuimov and Panchenko, 2023). The third combines network signals — how content propagates and who spreads it — with textual analysis (Ruffo et al., 2023; Lai, Toriumi and Yoshida, 2024). A parallel strand addresses machine-generated text: the MULTITuDE benchmark showed that detectors of AI-generated text degrade sharply across languages, and fine-tuned multilingual detectors have since become competitive (Madabushi et al., 2022; Macko et al., 2023; Spiegel and Macko, 2024).

The Indian case. The project webinar on misinformation detection in Indian languages, delivered by Asif Ekbal (IIT Patna) and Rina Kumari (KIIT, Bhubaneswar), illustrated both the difficulty and the ingenuity of low-resource detection (see Ekbal and Kumari, 2024). India has over 1,500 languages, 22 with official status, pervasive code-mixing, and minimal fact-checking infrastructure outside the largest languages. The presenting team's multitask architecture treats novelty detection and emotion recognition as auxiliary tasks alongside misinformation detection — exploiting the finding that false stories are engineered to appear novel and to trigger fear, disgust and surprise (Kumari et al., 2022), yielding improvements of roughly 7–28% over single-task baselines. Their subsequent multimodal work created a 10,473-item dataset in Hindi, Bengali and Tamil for detecting genuine images re-shared in false contexts, using multilingual encoders, contrastive learning and web-retrieved background knowledge. The team's stated research gaps (large-scale multilingual datasets, code-mixed benchmarks, robustness to generative AI) read as a research agenda for low-resource languages everywhere.

LLMs as friends. The project webinar “Multilingual Disinformation and Generative AI: Are LLMs Friends or Foes?” delivered by Carolina Scarton presented experiments showing that LLMs, for all their English-centricity, enable genuinely new support for extremely low-resource languages. For languages with scripts unseen in training, conventional fine-tuning largely fails; but zero-shot prompting augmented with simple lexical resources — word- or sentence-level alignments to a high-resource language — achieved usable classification accuracy in languages such as Tigrinya for which no dedicated tools exist. Because the method requires only a dictionary rather than a training corpus, it can in principle be redeployed to disinformation-relevant tasks (stance detection, persuasion-technique classification) in any minority language, providing timely signals for fact-checkers. Tooling built on such research is already deployed: fact-checking organisations, including Agence France-Presse (AFP) embed academic detection tools in browser plugins used by tens of thousands of practitioners.

LLMs as foes. The same webinar reported a large-scale evaluation of LLMs' capacity to generate personalised disinformation: eight models (three commercial, five open-weight) prompted over 66,000 times to produce false news articles localised by country, generation and political orientation, in English, Russian, Portuguese and Hindi. All models generated disinformation at high rates; the most permissive complied with essentially every request and produced the most convincingly personalised output, while even the most safety-conscious complied in nearly half of cases. Personalised outputs contained more named entities and stronger in-group affiliation language, which are precisely the features that make disinformation persuasive. Combined with the finding that safety filters are weakest in low-resource languages, the generation threat is asymmetric: easiest to aim at the communities least protected. The webinar's conclusion bears quoting in substance:

higher education institutions in the EU and UK have an obligation to sustain this safety research, because the companies will not, and comparable research capacity in the US is being dismantled.

3.3 Journalism and media

Fact-checking under strain. The project webinar delivered by Susan Daly of The Journal, Ireland's first online-only news outlet and its principal fact-checking operation, traced a decade's evolution. In 2016, the unit checked claims that politicians made on legacy broadcast media. Today, most claims originate on digital platforms, and sifting through them consumes more effort than checking them. Meanwhile, trust in fact-checking is itself under deliberate attack, and post-pandemic news avoidance is shrinking the audience for corrections. The unit's response has been structural: embedding verification skills across the newsroom, prioritising pre-bunking over the Sisyphean task of debunking, building a public knowledge bank of explainers, and treating audience questions as the primary signal of where friction and confusion lie.

The Anglophone conduit. Ireland's information ecosystem demonstrates that language, not geography, is the operative border online. US-origin narratives arrive intact and immediately; UK actors amplify Irish events as proof points in their own culture wars; and for years Irish fact-checkers saw claims that only later — after translation — reached continental European colleagues, a lag that EDMO-coordinated exchange reveals to be closing, plausibly because AI translation now moves narratives across linguistic boundaries as fast as they emerge. A study of X's Community Notes by the Spanish fact-checker Maldita found The Journal among the most-cited fact-checking sources globally in some periods — wildly disproportionate to its size and Irish focus, and a proxy measure of how underserved non-Anglophone communities are.

Minority-language media as counter-infrastructure. Minority-language media research (e.g., Manias-Muñoz et al., 2025) was represented in the project workshop by Craig Willis (European Centre for Minority Issues and chair of Plurilingmedia COST Action on minority-language media). Such research frames public service and community media in languages such as Irish, Welsh, Scottish Gaelic and Basque as resilience infrastructure: where professional media exist in a language, disinformation must compete with trusted, verifiable sources for that community's attention; where they do not, closed channels— WhatsApp, Telegram and similar messaging services — fill the vacuum. International evidence points the same way: US research finds ethnic and diaspora media effective counter-disinformation actors, while closed messaging spaces function as “information cocoons” for monolingual communities (ICFJ, 2025). In Ireland, The Journal's experiments are instructive: producing fact-checks in Irish for circulation in Gaeltacht communities, and translating explainer content into Polish for distribution through Polish Saturday schools, a one-off project that demonstrated demand the State does not currently meet. Furthermore, engagement with disinformation is itself culturally patterned: multilingual behavioural research finds that the sentiment–engagement relationship differs across language communities, so mitigation cannot be culture-blind (Viola, 2025).

Case: reaching Ireland's Polish-speaking community

During a public-interest information project, The Journal found that a substantial cohort of first-generation Polish residents in Ireland engaged with English-language media only functionally and had little awareness of Irish news or fact-checking sources. The outlet translated explainer content into Polish and distributed it through Polish-language Saturday schools, where teenagers carried it home to parents. The initiative worked and ended with the project, because no funding stream exists for sustained provision. Statutory public-information obligations in Ireland attach to Irish, but to no community language, however large its speaker base.

3.4 Political science and regulation

The Digital Services Act's promise and its language gap. The Digital Services Act (DSA) obliges very large online platforms to assess and mitigate systemic risks (Articles 34–35), report transparently on moderation, grant researchers data access (Article 40), and submit to audits; since its integration into the DSA framework in February 2025 (effective from July 2025), adherence to the strengthened Code of Conduct on Disinformation may count as an appropriate risk-mitigation measure for signatory platforms, and EDMO's network of national and regional hubs monitors disinformation and Code implementation across the Union. Participants in the project workshop, including Aidan O'Brien and Rodrigo Cetina, who evaluates platform compliance, were nonetheless blunt about the language gap: risk assessments seen to date rarely disaggregate by language; transparency reports mostly do not mention non-dominant languages at all; researcher data access is exercised overwhelmingly for English; and enforcement has yet to be triggered by a minority-language harm. The framework, in short, is sound and its implementation English-shaped (see also Dumbrava, 2023, on AI, democracy and elections).

The AI Act and the liability frontier. For generative models, the AI Act currently imposes chiefly transparency duties (users must know they are talking to a machine), which workshop participants judged insufficient for the disinformation problem. The more consequential developments may come through liability: participants pointed to emerging case law treating LLM output as the provider's own content rather than third-party content, and to Court of Justice jurisprudence under which a platform that actively shapes what users see — ranking, amplifying, recommending — may forfeit neutral-intermediary protections. Either line, if consolidated, would materially change platform incentives, including for the minority-language content they currently neglect. Workshop participants also stressed the complement to regulation: AI and media literacy, with current public scepticism toward AI seen as a rare opening for literacy efforts, while cautioning against measures such as blanket age verification whose effectiveness and rights implications remain unproven.

The Irish institutional gap. Workshop participants with direct knowledge of the Irish landscape described the State as far behind European peers, including smaller and less-resourced states, in counter-disinformation research infrastructure: no systematic monitoring of disinformation in Ireland exists in English, still less in Irish. Therefore, basic questions (which platforms carry the greatest risk, who is exposed, and where content originates) cannot currently be answered with evidence. Ireland hosts the European headquarters of many of the platforms concerned, giving its regulator, Coimisiún

na Meán, outsized relevance to DSA enforcement Union-wide. A National Counter Disinformation Strategy was published in April 2025; the workshop's assessment is that without a funded, sustainable research capability beneath it, the strategy has no evidence base on which to act.

3.5 Interdisciplinary perspectives

Complementary methods. The project workshop modelled the methodological pluralism the field requires. Actor-centred, ethnographic monitoring (who spreads disinformation and why) and narrative-centred comparative analysis (which claims recur across countries and languages) proved complementary: narratives repeat across borders, but their local carriers and targets differ, and both matter for response. Both approaches depend on platform data access, which has contracted sharply. Presenters on linguistic topics in the webinar series likewise reported that limited access prevents cross-platform analysis. This makes DSA Article 40 data access, extended to minority-language research, a genuine research-policy priority.

Closed messaging. Across the webinars and workshop, closed messaging applications recur as the least-observable vector: encrypted and semi-closed groups carried deepfakes to millions in the Indian election, army-lockdown rumours in Irish fuel protests, and health disinformation to migrant communities during COVID-19. Research methods here are underdeveloped everywhere, and effectively absent for minority languages; tiplines, community partnerships and voluntary data donation are the current frontier.

Participation and data sovereignty. A consistent normative thread runs from the webinars' analysis to their prescriptions, particularly in Rodrigo Cetina's presentation: communities that speak under-resourced languages should co-design the systems that will mediate their information environment — defining use cases, contributing and governing corpora, reviewing deployments and receiving benefit — under frameworks for collective data governance. Certification offers a practical mechanism: the Ecuadorian judiciary's approach (see box) conditions the use of AI in minority-language contexts on demonstrated, certified adequacy rather than assumed universality. The webinar series' closing formulation is a fitting summary of the interdisciplinary consensus: language technology inequality is not a technical problem but a question of justice, fairness and non-discrimination.

Case: Ecuador's moratorium on AI in minority-language court proceedings

Guidelines developed with the government of Ecuador for the use of AI across the court system include a moratorium on any AI use in cases involving speakers of Kichwa or Shuar, the country's co-official indigenous languages — not as a permanent prohibition, but as a certification condition: the moratorium lifts for any model demonstrated to perform as reliably in those languages as in Spanish. The guidelines pair this with principles of technological equality (tools must work equally well for everyone they are used on) and layered transparency (providers must disclose training-data limitations by language; courts must disclose when AI-informed decisions are made). The model translates directly to European debates: rather than banning AI from minority-language public services or silently accepting degraded performance, deployment should be conditional on certified adequacy.

Relevant recommendations: R1–R4, R8–R11.

4. Implications for academics and policymakers

For policymakers, the central implication is that language must become a unit of analysis in digital regulation. Every major instrument now in force (the DSA's risk-assessment and transparency machinery, the Code of Conduct on Disinformation, the AI Act's evaluation obligations, and FIMI monitoring) has the potential to protect minority-language communities. However, none currently requires reporting, resourcing or enforcement to be systematically disaggregated by language, leaving those communities without consistent protection. That is a correctable design choice, and the cost of correcting it falls largely on platforms already subject to the obligations concerned. For Ireland specifically, the priority ordering is clear: first build measurement (there is no policy without evidence of what circulates in Irish, English and community languages); then extend provision (public information and media literacy beyond English and Irish); then invest in Irish-language data infrastructure so that detection tools become possible at all. Ireland's position as host to the platforms' European headquarters gives it both particular responsibility and particular leverage.

For academics, three implications follow. First, the research gap is itself the finding: the volume and influence of dis/misinformation in Irish cannot presently be quantified because no one has systematically looked, and establishing that baseline is a prerequisite for every downstream question. Second, methodological realism is required: for languages like Irish, promising technical approaches include cross-lingual transfer, dictionary-augmented prompting and evidence-based verification, rather than bespoke models trained from scratch. These approaches are comparatively inexpensive to sustain and tend to degrade gracefully under resource constraints. Third, the field's dependence on platform goodwill for data is untenable; researchers should collectively invest in DSA Article 40 access, tiplines and community partnerships as durable alternatives. Across both audiences, the deepest implication is the one the webinar series ended on: whether AI systems serve all language communities or only the already-powerful is being decided now, in procurement choices, enforcement priorities and research funding — and defaults, once set, compound.

Relevant recommendations: R5–R10.

5. Future research agenda

Baseline measurement for Ireland. A systematic study of the volume, sources, vectors and targets of dis/misinformation circulating in Ireland, disaggregated across English, Irish and principal community languages, combining platform data (via DSA Article 40), fact-checker archives and community tiplines. This is the single most enabling project for Irish policy.

An Irish-language corpus and detection benchmark. An openly licensed, community-governed corpus of verified true and false claims and news content in Irish, supported by annotation guidelines adapted to Irish-specific deceptive strategies, would establish a public benchmark for dis/misinformation detection. Such a benchmark is a necessary foundation for the machine-learning model for minority languages, beginning with Irish, proposed in this project.

Cross-lingual transfer evaluation for Irish. Systematic evaluation of transfer-based dis/misinformation detection approaches (multilingual fine-tuning, dictionary-augmented zero-shot prompting, and cross-lingual evidence retrieval) on Irish, assessing how findings from other low-resource language settings generalise to a language characterised by high levels of bilingualism and substantial machine-translated content.

Machine-generated content in minority languages. Extension of machine-generated-text detection to Irish and comparable languages, including the characteristic artefacts of machine translation, given the documented use of automatic translation to include minority languages in influence operations and scams.

Closed-messaging methods. Development of ethically sound methods (tiplines, data donation, partnerships with community organisations) for studying dis/misinformation in closed messaging channels used by minority-language and migrant communities.

Regulatory auditing by language. A longitudinal audit of DSA risk assessments, transparency reports and Code of Conduct reporting for their treatment of non-dominant and non-official languages, producing an annual language-gap index to inform enforcement by the European Commission and Coimisiún na Meán.

Comparative minority-language resilience. Comparative study of how minority-language media ecosystems (Irish, Welsh, Basque, Catalan and others) confer resilience against disinformation, connecting media-policy variables to measured information quality, in collaboration with existing European networks on minority-language media.

Appendix: MULTIDISIN webinar series and workshop

Webinar 1 — No Language Left Behind? Disinformation, Digital Exclusion, and the Compounding Vulnerabilities of Under Resourced Languages in the Age of AI

Recording: https://youtu.be/YnE34Mu_32s

Speaker: - Rodrigo Cetina, Universitat Pompeu Fabra

Webinar 2 — Multilingual Disinformation and Generative AI: Are LLMs Friends or Foes?

Recording: https://youtu.be/Spe90r8ysw8

Speaker: Carolina Scarton, University of Sheffield

Webinar 3 — The Intersection of Fact-Checking and News Journalism in an Anglophone Ecosystem

Recording: https://youtu.be/5Nj4w2OYQ-E

Speaker: Susan Daly, The Journal.

Webinar 4 — Probing Misinformation Detection on the Shoulder of Affective Information and Multitasking

Recording: https://youtu.be/0GYu5qAhqKY

Speakers: Asif Ekbal, Indian Institute of Technology Patna; Rina Kumari, Kalinga Institute of Industrial Technology, Bhubaneswar.

Webinar 5 — Interdisciplinary Encounters with Disinformation: Lessons from the Fakespeak and NxtGenFake Projects

Recording not available

Speaker: Silje Alvestad, University of Oslo

Workshop Presentation — Platformed Hate Speech and Fake News on Social Media

Recording: https://youtu.be/yX2yVTjnTug

Speaker: Encarnación Hidalgo Tenorio, University of Granada

Workshop Panel Discussion: Interdisciplinary Approaches to Disinformation Research: Multilingual, Minority Language Perspectives and Beyond

Participants: Rodrigo Cetina, Universitat Pompeu Fabra; Carolina Scarton, University of Sheffield; Craig Willis, European Centre for Minority Issues & Chair of Plurilingmedia COST Action; Aidan O'Brien, European Digital Media Observatory Ireland & DCU FuJo

Bibliography

Ahuja, N. and Kumar, S. (2023) Mul-FaD: attention based detection of multiLingual fake news. Journal of Ambient Intelligence and Humanized Computing, 14, pp. 2481–2491.

Aira, T. and Cetina Presuel, R. (2024). Countering disinformation in minority and minoritized language communities: From bridging the digital language divide to context specific responses. UPF Barcelona School of Management.

Blasi, D., Anastasopoulos, A. and Neubig, G. (2022) Systematic inequalities in language technology performance across the world's languages. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics.

Bontcheva, K., Papadopoulous, S., Tsalakanidou, F., Gallotti, R., Dutkiewicz, L., Krack, N., Teyssou, D., Severio Nucci, F., Spangenberg, J., Srba, I., Aichroth, P., Cuccovillo, L., and Verdoliva, L.. (2024) Generative AI and disinformation: Recent advances, challenges, and opportunities. White paper, European Digital Media Observatory.

Chen, C. and Shu, K. (2024) Combating misinformation in the age of LLMs: opportunities and challenges. AI Magazine, 45(3).

Council of Europe (1992) European Charter for Regional or Minority Languages. Strasbourg.

Culloty, E. and Suiter, J. (2021) Disinformation and manipulation in digital media: Information pathologies. Routledge.

Dementieva, D., Kuimov, M. and Panchenko, A. (2023) Multiverse: multilingual evidence for fake news detection. Journal of Imaging, 9(4), 77.

Dumbrava, C. (2023) Artificial Intelligence, democracy and elections. Briefing, European Parliamentary Research Service.

Ekbal, A. and Kumari, R. (2024) Dive into misinformation detection: From unimodal to multimodal and multilingual misinformation detection. The Information Retrieval Series, vol. 30. Cham: Springer

Erdocia, I., Migge, B. and Schneider, B. (2024) Language is not a data set—Why overcoming ideologies of dataism is more important than ever in the age of AI. Journal of Sociolinguistics, 28, 20-25.

Erdocia, I., Schneider, B. and Migge, B. (2026) Language in the age of AI technology: From human to non-human authenticity, from public governance to privatised assemblages. Language in Society, 55(3), 523-543.

European Commission (2022) Regulation (EU) 2022/2065 on a single market for digital services (Digital Services Act), esp. Articles 34, 35 and 40.

European External Action Service (2024) Report on foreign information manipulation and interference (FIMI) threats. Brussels.

European Union (2024) Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act).

Francisco, M., Benítez-Castro, M. Á., Hidalgo-Tenorio, E., and Castro, J. L. (2022) A semi-supervised algorithm for detecting extremism propaganda diffusion on social media. Pragmatics and Society, 13(3), 532-554.

Government of Ireland (2025) National counter disinformation strategy. Dublin: Department of Culture, Communications and Sport (published April 2025).

Grieve, J. and Woodfield, H. (2023) The language of fake news. Cambridge: Cambridge University Press.

International Center for Journalists (2025) Disarming disinformation: United States case study — the role of ethnic and indigenous media. Washington, DC.

Kubś, J. (2025) Global offensive: Mapping the sources behind the Pravda network. Globsec.

Kumari, R., Ashok, N., Ghosal, T. and Ekbal, A. (2022) What the fake? Probing misinformation detection standing on the shoulder of novelty and emotion. Information Processing & Management, 59(1), 102740.

Lai, C., Toriumi, F. and Yoshida, M. (2024) A multilingual analysis of pro-Russian misinformation on Twitter during the Russian invasion of Ukraine. Scientific Reports, 14.

Macko, D., Moro, R., Uchendu, A., Lucas, J., Yamashita, M., Pikuliak, M., Srba, I., Le, T., Lee, D., Simko, J., and Bielikova, M. (2023, December). MULTITuDE: Large-scale multilingual machine-generated text detection benchmark. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 9960-9987).

Madabushi, H. T., Gow-Smith, E., Garcia, M., Scarton, C., Idiart, M. and Villavicencio, A. (2022, July). SemEval-2022 task 2: Multilingual idiomaticity detection and sentence embedding. In Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022) (pp. 107-121).

Manias-Muñoz, Bober, S. and Willis, C. (2025) Minority language media: Current challenges in a fragmented mediascape. Palgrave Macmillan.

Nicholas, G. and Bhatia, A. (2023) Lost in Translation: Large Language Models in Non-English Content Analysis. Washington, DC: Center for Democracy & Technology.

NLLB Team, Costa-jussà, M.R. et al. (2022) No Language Left Behind: scaling human-centered machine translation. arXiv preprint.

Põldvere, Nele, Elizaveta Kibisova, and Silje Susanne Alvestad. (2023) Investigating the language of fake news across cultures. In The Routledge handbook of discourse and disinformation, edited by Stefania Maci, Massimiliano Demata, Mark McGlashan, and Philip Seargeant, 153–165. London: Routledge. doi.org/10.4324/978100322 4495-11.

Rananga, S., Isong, B., Modupe, A., Isong, A. and Marivate, V. (2024) Misinformation detection : a review for high and low resource languages. Journal of Information Systems and Informatics, vol. 6, no. 4, pp. 2892-2922. DOI: 10.51519/journalisi.v6i4.931.

Ruffo, G., Semeraro, A., Giachanou, A. and Rosso, P. (2023) Studying fake news spreading, polarisation dynamics, and manipulation by bots: a tale of networks and language. Computer Science Review, 47.

Simon, F.M., Altay, S. and Mercier, H. (2023) Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. Harvard Kennedy School Misinformation Review, 4(5).

Spiegel, M. and Macko, D. (2024) KInIT at SemEval-2024 Task 8: fine-tuned LLMs for multilingual machine-generated text detection. Proceedings of SemEval-2024.

Viola, L. (2025) What about language? A multilingual behavioural study of user engagement with disinformation on X. Proceedings of ROMCIR 2025, CEUR Workshop Proceedings, Vol. 3986.

Vogel, I. and Meghana, M. (2020) Detecting fake news spreaders on Twitter from a multilingual perspective. 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), pp. 599–606.

Vosoughi, S., Roy, D. and Aral, S. (2018) The spread of true and false news online. Science, 359(6380), pp. 1146–1151.

Wang, X., Koneru, S., Venkit, P.N., Frischmann, B. and Rajtmajer, S. (2024) Monolingual and multilingual misinformation detection for low-resource languages: a comprehensive survey. arXiv preprint.

Yong, Z.-X., Menghini, C. and Bach, S.H. (2023) Low-resource languages jailbreak GPT-4. arXiv preprint.