AI-slands

AI and Island Cultures

Preservation, Extraction and the Right to Refuse

Andrew Jennings

Professor in Island Studies
Institute for Northern Studies
University of the Highlands and Islands

AI-slands Working Paper No. 2 | 2026

AI-slands research initiative: ai-slands.neohjalt.chatgpt.site

Suggested citation: Jennings, A. (2026) AI and Island Cultures: Preservation, Extraction and the Right to Refuse. AI-slands Working Paper No. 2. Institute for Northern Studies, University of the Highlands and Islands.

Abstract

Artificial intelligence is entering island cultural life through translation, speech recognition, archives, recommendation systems and generative tools. It can widen access to small-language media, reduce the labour of transcription and subtitling, and give island creators new means of production. Yet the same systems can misread culturally specific language, reward generic metropolitan aesthetics, extract cultural materials without meaningful consent and return synthetic versions of islands to the people from whom the source material came. This working paper argues that AI's cultural value for islands is conditional, not inherent. It should be judged by whether it increases the capacity of islanders to speak, remember, create and govern on their own terms. Evidence from Scottish Gaelic, Faroese, Kalaallisut, Sámi and Māori language technology shows both practical gains and recurring asymmetries in data, infrastructure and authority. Building on the AI-slands concepts of algorithmic mainlanding, synthetic islandness and insular AI sovereignty, the paper proposes an Island Cultural AI Charter and a research agenda for an AI-slands Observatory. The central conclusion is simple: preservation without authority can become extraction, and inclusion without the right to refuse is not cultural sovereignty.

Keywords: artificial intelligence; island cultures; language technology; cultural data; Indigenous data sovereignty; cultural governance

Evidence note

The references are limited to sources whose existence and bibliographical details were checked against publisher, proceedings or institutional repository records. Substantive claims were checked against the complete text or the relevant pages. DOI, repository and stable institutional links are supplied so that readers can inspect the evidence directly.

1. Culture is not simply content

Debate about artificial intelligence and islands often begins with access: can a tool translate the local language, search an archive, subtitle a recording or help a small cultural organisation reach a larger audience? Those are important questions, but they are not sufficient. Culture is not a stock of content waiting to be digitised. It is a living set of relationships among people, place, language, memory, practice and authority. The meanings of a story, song, place-name or customary practice do not reside only in its words or images. They also depend on who may tell it, in what season or setting, for which audience, and with what obligations.

This distinction matters particularly for islands. Island cultures are frequently rendered from outside through a narrow repertoire: remoteness, purity, decline, escape, wilderness or picturesque tradition. Artificial intelligence can accelerate that rendering because it learns from already circulating descriptions and images. The first AI-slands working paper called this process *algorithmic mainlanding*: systems trained and governed elsewhere treat metropolitan assumptions as normal and island difference as a deviation to be simplified (Jennings, 2026). Generative systems add a further risk of *synthetic islandness*, in which statistical familiarity produces plausible but culturally thin islands.

There is therefore no universal balance sheet of AI "pros and cons". The same capability can preserve and appropriate at once. Speech recognition can make an archive searchable while converting ancestral voices into reusable training data. Machine translation can open a Faroese article to new readers while mistranslating the concepts that make it Faroese. Image generation can lower costs for a local artist while flooding platforms with generic island imagery. The relevant question is not whether AI is good or bad for island culture. It is: **under whose authority does it operate, who benefits, and what remains possible to withhold?**

This paper develops that question through a northern and Arctic core—Scottish Gaelic, Faroese, Kalaallisut and Sámi—alongside a Māori comparison from Aotearoa New Zealand. The comparison is not intended to collapse different island histories or political statuses. In particular, Indigenous sovereignty has legal, territorial and collective dimensions that cannot be reduced to a general metaphor for local control. The cases are brought together because they show, with unusual clarity, how language technology is also a contest over cultural authority.

2. What AI can strengthen

AI can make a material contribution to cultural continuity. For archives, automated transcription and speaker recognition can turn thousands of hours of sound into searchable records. Optical character recognition can help recover printed and handwritten texts. Translation, text-to-speech and speech-to-text can support education, broadcasting, public services and everyday communication. Recommendation and discovery tools can help audiences find local music, film and writing. Generative systems can assist with drafting, editing and prototyping where cultural workers have little time or money.

The clearest benefits arise when tools address needs defined by speakers and cultural organisations. In Scottish Gaelic, Evans et al. (2022) developed automatic speech recognition using broadcast and other speech data, reporting a final word error rate of 26.30 per cent and identifying transcription, subtitling and accessibility as practical applications. The figure is not evidence that the language problem has been "solved"; it is evidence that sustained language-specific work can produce useful infrastructure. Human review remains essential, especially for names, dialects and culturally dense speech.

Across Sámi language technology, Moshagen et al. (2024) describe the GiellaLT infrastructure and a long-term model based on open tools, grammatical knowledge and collaboration with language communities. This matters because many commercially fashionable AI systems assume that more data and larger models are always better. For small languages, carefully built dictionaries, analysers, keyboards and spellcheckers may deliver more reliable public value than a fluent but opaque chatbot. Open infrastructure also gives language communities and institutions greater capacity to maintain tools when a commercial provider changes priorities.

Greenlandic–Danish machine translation illustrates a similar point. Kristensen-McLachlan and Nedergård (2024) assembled and tested a new benchmark for Kalaallisut. Their experiments showed that a relatively simple system could outperform the tested GPT-based approaches. More importantly, the authors documented the limits of the corpus: much of the material came from translated news, could contain Danish influence and did not necessarily

represent natural Kalaallisut across domains. A benchmark is useful precisely because it makes such limits visible and contestable.

These cases suggest three genuine benefits. First, AI can reduce routine labour around transcription, search and first-pass translation. Second, it can increase the digital presence of languages that platforms have neglected. Third, community-centred development can create skilled work and institutional capability within island and Indigenous organisations. But technology supports revitalisation only as part of a wider ecology of speakers, teachers, families, media, public services and cultural practice. Bird (2020) warns that treating Indigenous languages principally as data commodities can repeat the relations that endangered them. A tool may assist intergenerational transmission; it cannot substitute for it.

3. Inclusion can conceal new hierarchies

The appearance of a language in an AI interface is easily mistaken for equality. In reality, natural-language processing remains radically uneven. Joshi et al. (2020) found that the field's resources and research attention were concentrated in a small share of the world's languages. Multilingual systems may display hundreds of language names while their quality, safety and cultural competence vary sharply among them.

Faroese provides a precise example. Simonsen and Einarsson (2024) asked human evaluators to examine GPT-4 translations of Faroese news and blog texts. The system performed well on much general material but repeatedly struggled with idioms, named entities and cultural context. It confused Faroese terms with Icelandic, translated a stone-paved place for drying fish as a generic “fish place”, and performed less well on idiom-rich blogs than on news. The authors concluded that synthetic parallel text might supplement scarce training material, but it was not a substitute for human translation and Faroese-centred data.

The problem is not merely lower accuracy. A fluent error can be more culturally consequential than an obvious failure because it appears authoritative. It may replace a local category with the nearest dominant-language concept, erase a relationship encoded in kinship terminology or standardise a dialectal form out of existence. Moshagen et al. (2024) likewise caution that large language models can produce apparently fluent material in low-data languages that is misleading or nonsensical. The user may not have enough language proficiency to detect the problem, while the platform's polished interface supplies confidence it has not earned.

This is algorithmic mainlanding at the level of language. The dominant language supplies the training volume, categories and default pathways; the smaller language is fitted around them. Even a well-intentioned system can then privilege standard over dialect, written over oral practice, contemporary over historical usage, and the most digitised genres over the most culturally significant ones. The result may be technical inclusion accompanied by internal cultural narrowing.

Island language technology should therefore be assessed beyond a single aggregate score. Evaluation should be direction-specific—translation into a language may be much weaker than translation out of it—and should include names, place-knowledge, idiom, dialect, code-switching and culturally significant domains. Native speakers and cultural specialists must

be able to identify failures, change priorities and stop deployment. Equality of interface is not equality of capability, and capability without cultural evaluation is not trustworthiness.

4. Cultural production and synthetic islandness

For island artists, writers, musicians, filmmakers and heritage workers, generative AI can be genuinely useful. It can help a small organisation prototype an exhibition, create captions and transcripts, explore a visual concept or adapt publicity for different audiences. Used transparently and under the creator's direction, such tools may lower barriers created by small markets, limited staffing and distance from specialist services.

The risks are structural rather than simply aesthetic. The CULTAI Independent Expert Group (2025) reports that generative and recommender systems are reshaping cultural value chains, including authorship, ownership, visibility and remuneration. Recommendation and ranking systems can reinforce dominant aesthetics, while cultural data may be absorbed into AI development without adequate recognition or benefit sharing. For island creators, whose work is already interpreted through external markets and tourism imaginaries, these pressures are acute.

Synthetic islandness emerges when a system reproduces the signs of island culture without the relationships that give them meaning. A generated image may combine the wrong architecture, dress, vegetation and ritual while remaining visually persuasive. A chatbot may invent folklore or flatten contested history into a confident summary. A music generator may imitate the audible surface of a tradition without acknowledging performers or restrictions on use. Repetition then matters: synthetic material enters search results, educational resources and promotional campaigns, becoming new data from which later systems learn.

This is not an argument that only islanders may represent islands, nor that cultures should be frozen behind boundaries. Island cultures have always been relational, mobile and inventive. The issue is asymmetry. A distant platform can generate and distribute millions of representations at negligible marginal cost, while a local creator must invest time in correction, explanation and proof of authenticity. If platforms reward volume and familiarity, culturally specific work may be made less visible precisely because it does not resemble the system's learned island template.

A responsible cultural policy must therefore address discoverability as well as production. Public funders and cultural institutions should ask whether AI tools direct audiences towards identifiable island-made work, whether creators can opt out of model training, whether synthetic media is labelled, and whether correction routes are usable from small communities. Cultural diversity cannot be measured by the number of generated styles. It depends on the continued capacity of people and institutions to make, circulate and sustain work on their own terms.

5. When preservation becomes extraction

Digitisation is often described as preservation, but preservation answers only what is kept and whether it remains accessible. It does not answer who may use it, for what purpose, or who gains from derivative products. This gap becomes critical when archives, recordings,

dictionaries, photographs and community websites are treated as available training material.

The CARE Principles for Indigenous Data Governance—Collective Benefit, Authority to Control, Responsibility and Ethics—were developed to place people and purpose alongside technical approaches to data reuse (Carroll et al., 2020). They challenge the assumption that making data findable or accessible automatically makes its reuse legitimate. Lewis et al. (2020) similarly argue that Indigenous approaches to AI must begin with specific protocols and reciprocal relationships; there is no single Indigenous perspective that a universal checklist can capture.

Te Hiku Media offers a practical counter-model. Its Māori speech-recognition work is based on a long-held broadcasting archive, detailed data curation and close dialogue among language specialists and technical staff. Leoni et al. (2024) show that this expertise shapes the identification and repair of model failures rather than being added as cultural consultation after technical decisions have been made. Spano and Zhang (2025) document Te Hiku's refusal of a commercial offer for recordings and its use of a Kaitiakitanga licence intended to ensure that collaborative projects benefit Māori and that derived work remains subject to care obligations.

The case clarifies why consent to record, digitise or publish is not necessarily consent to train a model. AI reuse can change scale, audience, purpose and reversibility. It may enable voice cloning, synthetic text or commercial services that were impossible when material entered an archive. Some knowledge is collectively held; some is restricted by age, gender, season, role or kinship; some should not circulate outside its context. Local Contexts' Traditional Knowledge Labels demonstrate one way communities can express such conditions within digital systems, including rules for sacred, ceremonial or seasonally restricted material (Local Contexts, n.d.).

The lesson for islands is not that every cultural dataset must be closed. Openness can support research, education and language development, and many communities actively choose it. The lesson is that openness should be a governed choice, not an inference made by a crawler. Cultural data stewardship needs granular permissions, provenance, agreed benefits, continuing review and credible withdrawal mechanisms. Otherwise, the language of preservation can legitimise a new extraction: culture leaves as data and returns as a product.

6. Sovereignty includes the right to refuse

Participation is frequently presented as the ethical remedy for AI. Yet an invitation to participate within a project whose purpose, infrastructure and ownership have already been fixed offers limited authority. Meaningful governance includes the power to define the problem, select the technology, control the data, set evaluation criteria, redirect resources and say no.

Carlson and Worrell (2026) describe Indigenous refusal not as fear of technology but as an ethical and political judgement about when, how and on whose terms engagement should occur. Refusal may protect knowledge, resist surveillance or reject a relationship that offers visibility without authority. It can also be constructive: declining one platform or licence

may create space for community-owned infrastructure, a different partnership or a slower project.

Indigenous data sovereignty and general island cultural governance should remain analytically distinct. Not every island community is an Indigenous people, and local preference does not automatically carry the same collective rights. Nevertheless, institutions responsible for island heritage can learn from Indigenous-led practice: authority should travel with cultural data; governance should be relational rather than reduced to a one-time consent form; and technical access does not erase continuing obligations.

Within AI-slands, *insular AI sovereignty* does not mean that every island must build a foundation model or data centre. It means retaining meaningful capacity to choose dependencies and govern their terms (Jennings, 2026). For culture, that capacity includes local archives with clear rights information; procurement conditions for vendors; paid roles for language and cultural experts; public-interest evaluation; and the ability to restrict, correct or remove material. Sovereignty is practical when it changes who can make a consequential decision.

7. An Island Cultural AI Charter

The following charter is proposed as a starting point for island cultural organisations, public bodies, researchers, funders and technology partners. It is deliberately short enough to use at the beginning of a project. Each principle should be translated into locally specific rules, and Indigenous communities' own laws and protocols take precedence.

1. **Community authority.** Identify the people and institutions with legitimate cultural authority before data collection or system design. Consultation is not a substitute for decision-making power.
2. **Purpose and necessity.** State the cultural purpose, intended users and expected benefit. Do not use AI where a simpler method would be more reliable, accountable or respectful.
3. **Consent and provenance.** Record where every dataset came from, the basis on which it may be used, and whether permission covers model training and derivative outputs. Public accessibility alone is not sufficient consent.
4. **Cultural integrity in evaluation.** Test names, dialects, idioms, oral forms and culturally significant domains with appropriately paid speakers and cultural experts. Publish limitations in language users can understand.
5. **Protection for restricted knowledge.** Provide granular access rules for sacred, sensitive, seasonal or role-restricted material. Exclude it from training and generation unless the relevant authority explicitly decides otherwise.
6. **Attribution and shared benefit.** Credit and compensate cultural labour. Agree how models, datasets, revenue, skills and infrastructure will benefit the contributing community over time.
7. **Human cultural authority.** Make synthetic content identifiable. Provide reachable processes for correction, challenge and redress, and do not allow automated output to overrule recognised language or cultural expertise.

8. **Exit and refusal.** Preserve the right to decline participation, withdraw material where feasible and end a partnership without losing access to community data or essential cultural services.

The charter is not a certification badge. A project may satisfy a form while violating the relationship. Its value lies in moving questions of authority, benefit and refusal upstream, before cultural materials are collected and technical dependency becomes difficult to reverse.

8. Implications for an AI-slands Observatory

An AI-slands Observatory could turn these principles into comparative evidence. It should not rank islands by technological adoption. It should investigate where capability and authority sit, whose cultural labour is counted, and how systems change cultural visibility. Five linked strands would provide a practical programme.

First, a **language capability audit** would test common systems in both translation directions and across names, place-terms, idioms, dialects and public-service contexts. Second, an **island representation audit** would examine how search, recommendation and generative systems portray specific islands, including recurring stereotypes and invented cultural material. Third, a **cultural data register** would document provenance, permissions, custodianship and restrictions for datasets used in public or publicly funded projects. Fourth, a **cultural labour and benefit account** would track who performs transcription, annotation, correction and moderation, who is paid, and where intellectual property and revenue flow. Fifth, a **refusal and redress record** would document whether communities can stop a use, obtain a correction or recover their data.

This programme would make comparison possible without assuming that all islands should reach the same technological endpoint. A Faroese translation service, a Hebridean sound archive, a Kalaallisut benchmark and a Māori speech-recognition project may require different measures. The shared question is whether AI expands cultural agency or transfers it elsewhere.

9. Conclusion

Artificial intelligence can help island communities search archives, subtitle speech, develop language tools, reach audiences and experiment creatively. Those benefits are real. They are also contingent on patient linguistic work, durable institutions and people who can judge the cultural meaning of an output. The strongest cases are not stories of technology arriving to rescue a culture. They are stories of communities and language organisations directing technology towards purposes they already value.

The dangers are equally concrete. Uneven data can make fluent systems culturally unreliable. Recommendation systems can narrow visibility around dominant aesthetics. Generative tools can multiply synthetic islandness. Archives and public websites can be recast as raw material, separating cultural data from the authority and obligations that accompanied it. Preservation then becomes extraction; digital inclusion becomes another form of mainlanding.

The governing principle should be straightforward: AI is culturally beneficial when it increases the capacity of islanders to speak, remember, create and decide on their own terms. That requires provenance, community authority, culturally competent evaluation, shared benefit, correction and the right to refuse. The future of island culture will not be secured by representing more of it inside machines. It will depend on whether island communities retain meaningful power over the terms of representation.

References

Bird, S. (2020) 'Decolonising speech and language technology', in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 3504–3519. https://doi.org/10.18653/v1/2020.coling-main.313

Carlson, B. and Worrell, T. (2026) 'The ethics of refusal: Why the Indigenous "no" matters in the age of artificial intelligence', *Ethical Space*, 23(2/3). https://doi.org/10.21428/0af3f4c0.74c8c09e

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. and Hudson, M. (2020) 'The CARE Principles for Indigenous Data Governance', *Data Science Journal*, 19, 43, pp. 1–12. https://doi.org/10.5334/dsj-2020-043

CULTAI Independent Expert Group (2025) *Report of the Independent Expert Group on Artificial Intelligence and Culture*. Paris: UNESCO. https://www.unesco.org/en/articles/new-expert-report-explores-how-ai-transforming-culture

Evans, L., Lamb, W., Sinclair, M. and Alex, B. (2022) 'Developing automatic speech recognition for Scottish Gaelic', in *Proceedings of the 4th Celtic Language Technology Workshop within LREC2022*, pp. 110–120. https://aclanthology.org/2022.cltw-1.16/

Jennings, A. (2026) *AI-slands: Towards an Island-Centred Approach to Artificial Intelligence*. AI-slands Working Paper No. 1. Institute for Northern Studies, University of the Highlands and Islands. https://ai-slands.neohjalt.chatgpt.site/

Joshi, P., Santy, S., Budhiraja, A., Bali, K. and Choudhury, M. (2020) 'The state and fate of linguistic diversity and inclusion in the NLP world', in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6282–6293. https://doi.org/10.18653/v1/2020.acl-main.560

Kristensen-McLachlan, R.D. and Nedergård, J.S.K. (2024) 'A new benchmark for Kalaallisut–Danish neural machine translation', in *Proceedings of the Fourth Workshop on NLP for Indigenous Languages of the Americas*, pp. 50–55. https://doi.org/10.18653/v1/2024.americasnlp-1.7

Leoni, G., Steven, L., Thompson, M., Keith, T., Mahelona, K., Jones, P.-L. and Duncan, S. (2024) 'Solving failure modes in the creation of trustworthy language technologies', in *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024*, pp. 325–330. https://aclanthology.org/2024.sigul-1.39/

Lewis, J.E. et al. (2020) *Indigenous Protocol and Artificial Intelligence Position Paper*. Honolulu, HI: Indigenous Protocol and Artificial Intelligence Working Group. https://doi.org/10.11573/spectrum.library.concordia.ca.00986506

Local Contexts (n.d.) 'Traditional Knowledge Labels'. https://localcontexts.org/labels/traditional-knowledge-labels/

Moshagen, S.N., Antonsen, L., Wiechetek, L. and Trosterud, T. (2024) 'Indigenous language technology in the age of machine learning', *Acta Borealia*, 41(2), pp. 102–116. https://doi.org/10.1080/08003831.2024.2410124

Simonsen, A. and Einarsson, H. (2024) 'A human perspective on GPT-4 translations: Analysing Faroese to English news and blog text translations', in *Proceedings of the 25th Annual Conference of the European Association for Machine Translation*, Vol. 1, pp. 24–36. https://aclanthology.org/2024.eamt-1.7/

Spano, I. and Zhang, Y. (2025) 'Indigenous data sovereignty in intangible cultural heritage governance: A complementary approach to public–private partnerships', *International Journal of Cultural Property*, 32(2), pp. 167–193. https://doi.org/10.1017/S0940739125100064