Media & Press All Europe Countries Artificial Intelligence

AI is making journalistic language more repetitive and predictable – and it's a problem for all of us

As AI tools increasingly generate news content, journalistic language is growing dangerously repetitive and predictable. This homogenization threatens critical thinking, flattens unique cultural narratives, and compromises the core diversity of public information.

Xosé López-García, Universidade de Santiago de Compostela and Cristian Augusto Gonzalez Arias, Pontificia Universidad Catolica de Valparaiso; Universidade de Santiago de Compostela

What happens to language when a growing amount of text published in the press, online and on social media is written by machines? This question is not just important for the profession of journalism – it also has an impact on the richness of the language we all use to comprehend, describe and discuss reality itself.

Historically, the press has been a space where public language grows and becomes richer. It is not, of course, the only driver of linguistic change, but it is one of the fields where new or emerging words, turns of phrase and ways of describing facts begin to circulate within society.

Studies on journalistic language and neologisms clearly demonstrate that newspapers are platforms for the creation and dissemination of new vocabulary, especially when it is needed

Co-funded by the European Union

to report on events, technology and social changes for a broad audience.

However, if a significant amount of journalistic writing is delegated to generative AI, this role will diminish. Large language models (LLMs) generally work by predicting the next "token" or word in a sequence. This allows them to produce fluent and believable text, but it also gives them a tendency to prioritise statistical regularity, as well as common, established arguments and formulations.

In and of itself, this does not degrade language. The problem arises when this logic comes to dominate writing in the public sphere.

The AI feedback loop

The risks become serious when AI systems begin training themselves on texts already produced by AI. This leads to what a number of studies call "model collapse", a degenerative process whereby material produced by one model contaminates the training data of later generations.

In plain terms, this means that AI systems learn more and more from synthetic text. If these texts fill public spaces – both online and offline – the verbal ecosystem for future training will be much more constricted.

A greater volume of artificial text means less contact with the social variation that is intrinsic to human language. It may well lead to linguistic decline in multiple fields.

It also entrenches existing biases and prejudices. When data variation drops and established patterns predominate, the biases in training material can be reinforced instead of being corrected. Research on the evolution of LLM bias warns that recursive processes can magnify existing prejudices instead of broadening perspectives.

In addition, writing is also becoming more repetitive and homogenised. It repeats syntactic structures, and tends towards a neutral tone, formulaic expressions, and predictable paragraph structures. This is especially important in journalism because the press does not just exist to broadcast information – it also mediates between specialised and more accessible registers, decides where to place emphasis, translates vocabulary, and teaches forms of expression.

When public language becomes too uniform, it limits

journalism's ability to fine tune writing in response to new information.

Eroding linguistic innovation

All of this leads to a reduction in the number of unusual or specialised words, less common constructions, and pragmatic nuance – a term that refers to devices such as irony, ambiguity and variation in points of view. The increasing use of synthetic text in AI training is also associated with a decline in performance and more limited coverage of human language diversity. Put simply, the system preserves the centre better than the fringes.

But in language, many innovations begin as erratic detours, unlikely word uses, or localised ways of naming a new phenomenon. If the system always favours the most statistically likely option, it means there is less space for emerging language to circulate and take root.

This point should not be understood as some abstract dichotomy between human and machine, but as a concrete difference – between language that is exposed to the chance events of human society, and textual output derived from pre-learned regularity.

Deteriorating public language ecosystems

This is not simply a question of having fewer distinct words, but also a reduced capacity to make subtle distinctions. When language becomes vaguer, or more predictable or repetitive, it also impoverishes the tools we use as a society to describe problems, clarify opinions and engage in public debate.

On a broader level, the problem is not confined to what happens to the AI models trained on this data, but also what happens to the public language ecosystem. If the internet becomes filled with synthetic text, readers, journalists and institutions will all be exposed to less diverse public language.

Some research also speaks of synthetic text "polluting" the online ecosystem, showing that the way we mix real data with artificial data is vital to preventing further decline.

All is not lost

Having said all this, we should not get carried away. Research does not show that all use of AI invariably leads to collapse or decline. Some studies show that when synthetic data is mixed

with real data rather than replacing it entirely, the collapse does not behave in the same way and the error can be contained.

In other words, the problem does not lie in occasionally using AI, or in the judicious combination of synthetic and human data. It occurs when human writing is replaced en masse, and when its replacement is then repurposed as if it were living language.

As AI becomes part of journalists' working lives, journalism is becoming more efficient. But what does a society lose when the language circulating in the public sphere becomes more uniform and predictable, and less open to innovation?

If the press gives up, even partially, its role of writing, translating, naming and teaching new language, it will not just affect journalists' workdays. It will also weaken one of the spaces where public language has most been able to enrich, renew and expand itself.

Xosé López-García, Periodismo digital, comunicación digital, Universidade de Santiago de Compostela and Cristian Augusto Gonzalez Arias, Investigador, Pontificia Universidad Catolica de Valparaiso; Universidade de Santiago de Compostela

This article is republished from The Conversation under a Creative Commons license. Read the original article.