intensity (in terms of the amount of timber extracted) below which species can persist even in the face of forest degradation[7].

Ewers *et al*. examined the responses of hundreds of species to varying intensities of selective logging on the island of Borneo in the Malaysian state of Sabah. The authors identified logging thresholds for conservation measures that are proactive (protection) and reactive (restoration). This study is a notable advance for at least three reasons.

First, the sheer breadth of the coverage of the tree of life is remarkable — the authors examined more than 1,500 species (or equivalent taxonomic units) of plants and animals varying greatly in key characteristics that determine responses to forest degradation, such as size, position in the food web and ecological function. Second, the scale of sampling is unparalleled in terms of understanding the effects of forest degradation on biodiversity — 127 surveys across 11 years, with logging intensity varying from 0% to 99% of timber extracted. Third, Ewers *et al*. identify two thresholds of timber extraction — a 'change point' at which there is an observable change in the probability of the occurrence of a species, and a 'maximum rate point' at which the rate of change in probability of occurrence is most rapid.

The authors also assessed the relationship between logging intensities and prevalence of particular species. For example, bird species that declined with greater logging intensity included the rhinoceros hornbill (*Buceros rhinoceros*; Fig. 1) and the argus pheasant (*Argusianus argus*).

The data Ewers and colleagues present have implications for conservation management. Crucially, the authors find that even minimal logging causes species losses, reinforcing the importance of primary forests for preserving the full complement of tropical biodiversity[3]. Furthermore, forests that have lost less than the 'change point' (roughly 30% of timber mass) retain high levels of biodiversity and ecological function, making it worthwhile to protect them, taking a proactive conservation approach. Over time, natural regeneration in such forests should lead to even better conservation outcomes without the need for expensive active restoration.

At 70% timber-biomass loss, the 'maximum rate point', most species showed fast declines in occurrence. Bringing biodiversity and ecosystem function back to such forests will probably require targeted and assisted restoration measures, such as planting native trees and removing invasive species; in other words, reactive conservation strategies. Because the rate of reduction in biodiversity is greatest at 70% of timber biomass loss, even small attempts to reverse degradation should lead to valuable biodiversity gains. Forests between 30% and 68% of timber biomass loss might need some degree of active intervention.

Many of the other patterns in Ewers and colleagues' study mirror known effects of selective logging — for instance, larger species and dietary or habitat specialist species are disproportionately negatively affected compared with small and generalist species[7].

Although the potential of the authors' findings to inform conservation policy for tropical biodiversity is immense, several questions remain. Whether these 30% and 68% thresholds apply to tropical forests and their biodiversity across the world needs further investigation. Although much more exhaustive in scope and coverage than any similar work, the data presented by Ewers and colleagues all come from a single location — from the Stability of Altered Forest Ecosystems (SAFE) Project. Bornean flora and fauna might respond to forest degradation in a different way from Amazonian or African species, given differences in evolutionary history and contemporary environments[8].

The continuing climate crisis adds further uncertainty. The species composition of ecological communities even in primary forests has been altered by climate change, shifting the baseline against which degraded forest communities are compared[9]. Furthermore, interactions between climate change and forest degradation could mean that the same species respond differently to climate change in primary and degraded forests[10]. Whether currently observed patterns will hold in the future despite the effects of climate change on biodiversity remains to be seen.

Forest degradation, in most cases, does not operate alone. It is often accompanied by other threats to biodiversity, such as the proliferation of roads (causing habitat fragmentation), hunting, fire and invasion by non-native species[11]. The precise mix of the intensities of these concomitant threats will vary from site to site, adding further complexity and uncertainty for biodiversity and conservation outcomes. Finally, biodiversity or ecological value is one of a variety of potentially conflicting considerations for conservation, including economics, cultural considerations, politics and social justice. Whether the recommendations of Ewers *et al*. are implemented will be governed by a multitude of factors.

Umesh Srinivasan is at the Centre for Ecological Sciences, Indian Institute of Science, Bengaluru 560012, India.
e-mail: umeshs@iisc.ac.in

1. Ewers, R. M. *et al*. *Nature* **631**, 808–813 (2024).
2. Laurance, W. F. *et al*. *Nature* **489**, 290–294 (2012).
3. Gibson, L. *et al*. *Nature* **478**, 378–381 (2011).
4. Pimm, S. L. & Raven, P. *Nature* **403**, 843–845 (2000).
5. Wright, S. J. *Trends Ecol. Evol.* **20**, 553–560 (2005).
6. Edwards, D. P., Tobias, J. A., Sheil, D., Meijaard, E. & Laurance, W. F. *Trends Ecol. Evol.* **29**, 511–520 (2014).
7. Burivalova, Z., Şekercioğlu, Ç. H. & Koh, L. P. *Curr. Biol.* **24**, 1893–1898 (2014).
8. Nowakowski, A. J. *et al*. *Ecol. Lett.* **21**, 345–355 (2018).
9. Stouffer, P. C. *et al*. *Ecol. Lett.* **24**, 186–195 (2021).
10. Williams, J. J., Freeman, R., Spooner, F. & Newbold, T. *Glob. Change Biol.* **28**, 797–815 (2022).
11. Lapola, D. M. *et al*. *Science* **379**, eabp8622 (2023).

The author declares no competing interests.
This article was published online on 17 July 2024.

Machine learning

AI returns gibberish when trained on generated data

Emily Wenger

Generative AI models are now widely accessible, enabling everyone to create their own machine-made something. But these models can collapse if their training data sets contain too much AI-generated content. **See p.755**

As generative artificial intelligence (AI) models — from Open AI's ChatGPT to Meta's Llama and beyond — become more available, the amount of AI-generated content on the Internet is swelling. AI-generated blogs, images and other content[1] are now commonplace (see go.nature.com/3yd2czz). And although the effects of an AI-generated Internet on humans remain to be seen, on page 755, Shumailov *et al*.[2] report that the proliferation of AI-generated content online could be devastating to the models themselves.

Conventional generative AI models learn to create realistic content by extracting statistical patterns from large swathes of Internet data — terabytes of articles, chat forums, blog posts and images. But what happens to the models if those chat forums and blog posts are AI-generated, as is increasingly the case? Shumailov *et al*. showed that large language

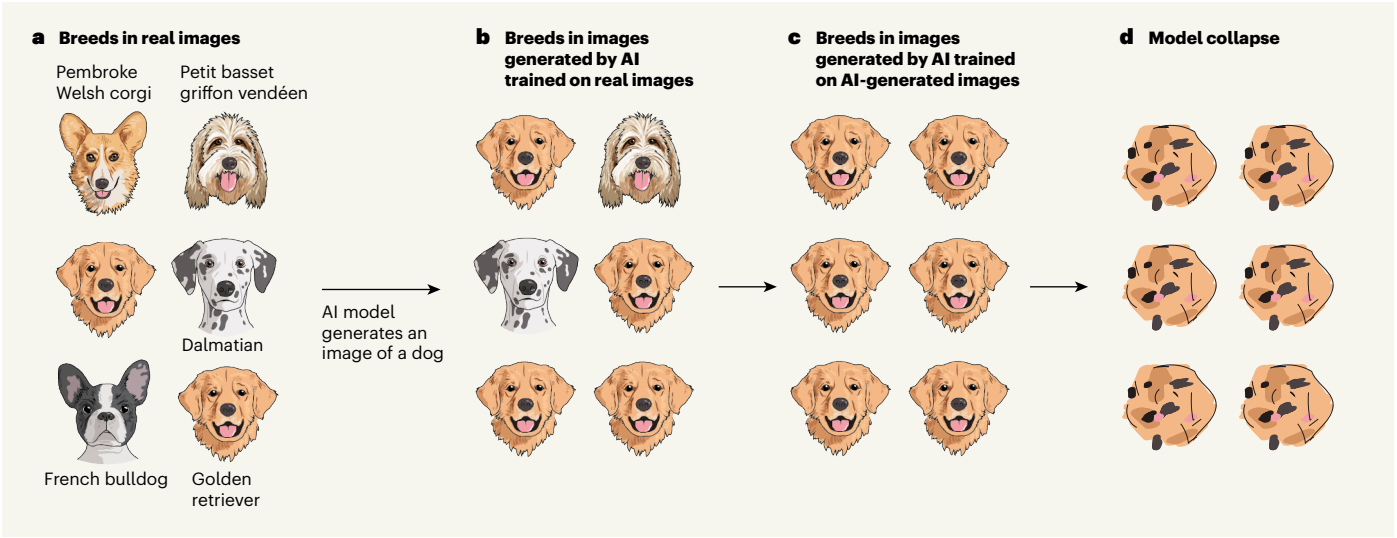

Figure 1 | Training an artificial intelligence (AI) model on its own output. **a**, An AI model will generate an image of a dog by learning from sets of real images, in which common dog breeds, such as golden retrievers, are over-represented, and rarer breeds, such as French bulldogs, Dalmatians, Pembroke Welsh corgis and petit basset griffon vendéens, are under-represented. **b**, The output of the model will therefore be more likely to resemble a golden retriever than a rarer breed. **c**, If the model is then trained on its own generated output, it might forget the more obscure dog breeds. Shumailov *et al.*[2] found that this is a general principle in the large-language-model setting; after several cycles of training the models on their own generated data, AI models eventually generate nonsensical outputs (**d**).

models (LLMs) 'collapse' when trained on their own generated content: after several cycles of outputting content and then being trained on it, the models produce nonsense.

This model collapse occurs because training models on their own generated content causes them to 'forget' the less-common elements of their original training data set (Fig. 1). Imagine a generative-AI model tasked with generating images of dogs. The AI model will gravitate towards recreating the breeds of dog most common in its training data, so might over-represent the golden retriever compared with the petit basset griffon vendéen, given the relative prevalence of the two breeds. If subsequent models are trained on an AI-generated data set that over-represents golden retrievers, the problem is compounded. With enough cycles of over-represented golden retrievers, the model will forget that obscure dog breeds such as petit basset griffon vendéens exist and generate pictures of just golden retrievers. Eventually, the model will collapse, rendering it unable to generate meaningful content.

Although a world overpopulated with golden retrievers doesn't sound too bad, consider how this problem generalizes to the text-generation models examined by Shumailov and colleagues. When AI-generated content is included in data sets that are used to train models, these models learn to generate well-known concepts, phrases and tones more readily than they do less-common ideas and ways of writing. This is the problem at the heart of model collapse.

Among other things, model collapse poses challenges for fairness in generative AI. Collapsed models overlook less-common elements from their training data, and so fail to reflect the complexity and nuance of the world. This presents a risk that minority groups or viewpoints will be less represented, or potentially erased. As the authors recognize, concepts or phrases that seldom feature in LLM training data are often the ones that are most relevant to marginalized groups. Ensuring that LLMs can model them is essential to obtaining fair predictions — which will become more important as generative AI models become more prevalent in everyday life.

So how can this problem be mitigated? Shumailov *et al.* discuss the possibility of using watermarks — invisible but easily detectable signals that are embedded in generated content — to enable easy identification and removal of AI-generated content from training data sets. Many generative-AI watermarks have been proposed and are used by commercial model providers such as Meta, Google and OpenAI.

Unfortunately, watermarks are not a panacea. Researchers have found that watermarks can be easily removed from AI-generated images[3]. Sharing watermark information also requires considerable coordination between AI companies, which might not be practical or commercially viable. Such coordination efforts suffer from a sort of prisoner's dilemma: if company A withholds information about its watermarks, its generated content could be used to train company B's model, resulting in B's failure and A's success. Other model providers could also simply choose not to watermark the output of their models.

Although Shumailov *et al.* studied model collapse in text-generation models, future work should investigate this phenomenon in other generative models, including multimodal models (which can produce images, text and audio) such as GPT-4o. Furthermore, the authors did not consider what happens when models are trained on data generated by other models, rather, they focused on the results of a model trained on its own output. Given that the Internet is populated by data produced by many models, the multi-model scenario is more realistic — albeit more complicated. Whether a model collapses when trained on other models' output remains to be seen. If so, the next challenge will be to determine the mechanism through which the collapse occurs.

As Shumailov *et al.* note, one key implication of model collapse is that there is a 'first-mover' advantage in building generative-AI models. The companies that sourced training data from the pre-AI Internet might have models that better represent the real world. It will be interesting to see how this plays out, as more companies race to make their mark in the generative-AI space — and, in doing so, populate the Internet with increasing amounts of AI-produced content.

Emily Wenger is in the Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708, USA and at Meta AI, New York, New York, USA.
e-mail: emily.wenger@duke.edu

1. Thompson, B., Dhaliwal, M. P., Frisch, P., Domhan, T. & Federico, M. Preprint at arXiv https://doi.org/10.48550/arXiv.2401.05749 (2024).
2. Shumailov, I. *et al. Nature* **631**, 755–759 (2024).
3. Zhao, X. *et al.* Preprint at arXiv https://doi.org/10.48550/arXiv.2306.01953 (2023).

The author declares no competing interests.