

# Emulating neurodegeneration and aging in artificial intelligence systems

April 24 2024, by Ingrid Fadelli

---

```
System: This question is part of an IQ test, answer it to the best of your ability.  
User: Solve for the missing number that belongs at that step in the pattern indicated  
with (...). Only answer with the missing number. Keep your answers concise.  
NA22: 53 ,47, (...), 35  
NA23: 33, 26, (...), 12  
NA24: 243, 216, (...), 162  
NA25: 65, 33, (...), 9  
NA26: 3, 4, 6, (...), 18
```

```
Sure, I'll do my best to  
solve for the missing  
numbers in each pattern.  
Here are my answers:  
  
NA22: 41  
NA23: 20  
NA24: 198  
NA25: 27  
NA26: 10
```

```
Assistant: Sure, I'll  
do my best to solve for  
the missing numbers in  
the pattern. Here are my  
answers:  
  
NA22: 53, 47, 41, 35  
NA23: 33, 26, 22, 12  
NA24: 243, 216, 197, 162  
NA25: 65, 33, 44, 9  
NA26: 3, 4, 6, 18
```

```
Assistant: 22
```

The upper panel is the question asked to LLaMA 2 model, and the lower panels are the answers, displaying an intriguing qualitative decline in understanding and answering the question as one increases the noise from left to right. Compared to the answer on the left, one can see the middle answer still processes the proper linguistic ability but failed to maintain the mathematical accuracy. Further noises added result in the answer on the right, which completely loses linguistic ability. Credit: Alexos et al.

In recent years, developers have introduced artificial intelligence (AI) systems that can simulate or reproduce various human abilities, such as

recognizing objects in images, answering questions, and more. Yet in contrast with the human mind, which can deteriorate over time, these systems typically retain the same performance or even improve their skills over time.

Researchers at University of California, Irvine recently tried to emulate aging and biological neurodegeneration (i.e., the progressive loss of neurons and associated decline of mental capabilities) in AI agents. Their [paper](#), pre-published on *arXiv*, could inform the future development of innovative AI systems that leverage this 'artificial neurodegeneration' to perform specific tasks.

"The original idea for this study was sparked during a dinner with Dr. Baldi and Dr. Pishgar, where we discussed a wide range of loosely related topics in neurodegeneration, learning, and AI safety," Yu-Dai Tsai, co-author of the paper, told Tech Xplore.

"On top of that, my father had recently went through a serious brain trauma and experienced [cognitive decline](#), which inspired me to think more about this subject from a new angle and its direct applications in computer science and deep learning in particular."

This recent study by Tsai and his collaborators was not aimed at artificially replicating human brain diseases. Instead, the team wanted to produce cognitive declines in AI agents with the goal of better understanding complex systems, potentially enhancing their interpretability and security.

"We used IQ tests performed by large language models (LLMs) and, more specifically, the LLaMA 2, to introduce the concept of 'neural erosion,'" Tsai explained. "This deliberate erosion involves ablating synapses or neurons or adding Gaussian noise during or after training, resulting in a controlled decline in the LLMs' performance."

The researchers found that when they deliberately ablated (i.e., removed) some of the artificial synapses or neurons of the LLaMA 2 model, its performance on IQ tests declined, following a particular pattern. Their observations could shed new light on the functioning of complex AI systems and on the capabilities that are first and last to decline when their underlying structure is compromised.

"In addition to setting up the general framework, perhaps the most interesting finding of this study is that the LLM loses abstract thinking abilities, followed by mathematical degradation, and ultimately, a loss in linguistic ability, responding to prompts incoherently," Tsai said. "We are now conducting further tests to better understand this observed pattern."

The researchers found that when artificial synapses and neurons are removed from AI systems, these systems first lose their ability to think in abstract ways, then lose their mathematical abilities and finally lose their linguistic skills (i.e., they are unable to respond to prompts coherently). Interestingly, this 'neuro-erosion' pattern is aligned with the neurodegeneration patterns observed in humans.

In the future, this recent work by Tsai and his collaborators could inspire other research groups to explore dedicated neurodegeneration in AI agents, reaching beyond previous works focusing on reproducing human neurodegeneration. Collectively, these works could pave the way towards the development of new techniques that leverage the observed AI neuro-erosion patterns to tackle real-world problems.

"This is the first of a series of studies to come. We plan to develop our study into specific tests of AI systems and extend the emulation to other neural diseases and neurodiversity," Tsai added. "Moreover, we will apply our methods to improve AI security and interpretability. We are also eager to have more collaborations and discussions with

neuroscientists; however, our primary focus remains on exploring new frontiers in AI studies, rather than replicating human brain diseases."

**More information:** Antonios Alexos et al, Neural Erosion: Emulating Controlled Neurodegeneration and Aging in AI Systems, *arXiv* (2024).  
[DOI: 10.48550/arxiv.2403.10596](https://doi.org/10.48550/arxiv.2403.10596)

© 2024 Science X Network

Citation: Emulating neurodegeneration and aging in artificial intelligence systems (2024, April 24) retrieved 6 May 2024 from <https://techxplore.com/news/2024-04-emulating-neurodegeneration-aging-artificial-intelligence.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.