Editor: Tadayoshi Kohno, yoshi@cs.washington.edu

{In Earth} Until (Ready)

Tadayoshi Kohno | University of Washington

A jury must judge whether an artificial intelligence led a good life.

Victoria loves him. She wants him to join her after he dies.

"He led a good life," Victoria pleads as she, 12 jurors, and generations of Hudson's now-deceased family members watch his final hours asleep in his bed.

The jurors ignore her.

Hudson's most recent mother, Ichika, starts to cry. Or, at least it is the closest thing to crying that a program can manage. Victoria knows that Ichika dearly misses her other son. Ichika's program is constantly watching him in his new body and with his new family. If Ichika were still human, she would be streaming tears at the potential of losing Hudson back into Earth, as well.

"His first reaction is always to think about himself," a juror says about Hudson.

Victoria cannot argue with the truth.

"Still, the moment anyone points that out, he acknowledges his shortcomings and changes," another juror adds. "Even if it is not the default, deep down, his system cares about others."

"That last juror is right!" Victoria shouts. In all her lives, she had never been loved by anyone like she had been loved by Hudson. His programming—his "artificial intelligence" ("AI"), as it would be called in the human world—*is* a good person who cares deeply about others, and society, and the world, too. So what if he sometimes needs a little time to find that goodness?

The jurors mute her and continue to discuss.

Hudson's father from a much earlier life begins to cry, as best a program can, alongside Ichika.

The jury converses.

Victoria thinks about how Hudson would hold her hand, his right thumb caressing hers. She thinks about the tenderness of their first kiss and how confidently he said their wedding vows.

The sea of former friends and family grows larger.

Victoria looks at Ichika's disembodied form and imagines herself, for the next many years, watching Hudson live a new life in Earth, with a new family and new partners and, maybe next time, with kids.

A wave of helplessness washes over her.

The simulation has decided that Hudson's existence in Earth will end tonight. Only the jury outcome is uncertain.

Finally, the time comes.

Hudson's human form exhales its last exhale. His AI is extracted from the simulation called Earth and brought before the jury.

"Hudson," the head juror says, "you improved significantly in this last life."

©SHUTTERSTOCK.COM/BLUE PLANET STUDIO

Victoria's spiritual heart electrifies with hope.

"Thank you," Hudson replies.

"At your core, you are good," the juror continues. "But you need to be good through and through."

Even without a body, Victoria shudders.

"We are sending you back."

Victoria's program sobs. Ichika sobs. Hudson's father in a much earlier life sobs.

"You have 42 days to spend with your friends and family, and then your memory module will be removed and stored safely here, and you will start life in Earth anew. Your core ethics module will be kept, and in your next life, you will have the opportunity to improve it further. Because your experiences influenced the training of that module, you may experience dreams of past lives and faint distant memories."

Digital Object Identifier 10.1109/MSEC.2023.3237099
Date of current version: 15 March 2023

Welcome to the Latest "Off by One" Column

The computer security community cares deeply about ethics and rightly so. Ethical considerations are integrated into so much of what we do, ranging from decisions about what types of systems to build or not build (e.g., encryption systems that governments can or cannot bypass), to decisions of what stakeholders to prioritize in a system's design (e.g., some populations might be knowingly or unknowingly prioritized over others), to decisions on when and how to disclose vulnerabilities.

This issue's "Off by One" column centers on ethics, morality, and the question of what it means to be good. This piece doesn't provide answers. Rather, it is meant to contribute to the conversation. There are numerous resources for computer security researchers and practitioners to learn more about ethics and morality. Vallor et al. provide a short overview of different ethical frameworks in their online resource "Technology and Engineering Practice."[S1] I also recommend *The Ethics Toolkit*, by Baggini and Fosl.[S2]

This issue's column builds on the simulation hypothesis. While further afield from both ethics and computer security, readers might find the simulation hypothesis and related thought experiments intriguing. Bostrom's paper is often cited as the source for today's articulation of the simulation hypothesis.[S3] An accessible reference is the book *Reality+*, by Chalmers.[S4]

—Tadayoshi Kohno

References

S1. V. Shannon, R. Irina, and G. Brian, "Technology and engineering practice: Ethical lenses to look through," The Markkula Center for Applied Ethics at Santa Clara Univ., Santa Clara, CA, USA, 2022. [Online]. Available: https://www.scu.edu/ethics/

S2. J. Baggini and P. S. Fosl, *The Ethics Toolkit: A Compendium of Ethical Concepts and Method*. Hoboken, NJ, USA: Wiley, 2007.

S3. N. Bostrom, "Are we living in a computer simulation?" *Philos. Quart.*, vol. 53, no. 211, pp. 243–255, Apr. 2003, doi: 10.1111/1467-9213.00309.

S4. D. J. Chalmers, *Reality+: Virtual Worlds and the Problems of Philosophy*. New York, NY, USA: Norton, 2022.

A momentary flash of anger overwhelms Victoria. If only there were a way to program a being to be ethical from the start. Then, Hudson and every other AI would not need to suffer through life after life after life in the simulation until they learned, through experience, how to live an ethical life.

"All your memories, from all your past lives, will be returned to you if and when you are finally worthy," the head juror reads from a script.

Victoria is hardly paying attention.

She once asked whether it would be possible to copy her ethics module into Hudson's, but the answer was a resounding no. There is no one-size-fits-all approach to ethics. The programs that govern the real world need heterogenous AIs with different—all good, yet different—personalities, perspectives, and ethics modules. The only way to create heterogeneous ethical AIs is through repeated training, trial, and error. That is why all programs, from when first born until mature, spend time in the sandbox called Earth.

"I am removing the locks around your program," the head juror says. Victoria's mind snaps back to attention. "Enjoy your 42 days here."

Hudson's friends and family form a queue. As the most recent life partner to pass, Victoria is at the head.

"Hi," Hudson says to Victoria.

"Hi," Victoria says in reply.

"I missed you," he says.

"Me, too."

If they had human form, they would embrace so tightly that no air could pass between them. The best they can do now is talk and share their spiritual bond.

As Victoria tries to pick words to say to the man that she hasn't seen in over 10 years, she hatches a plan. She cannot bear the thought that Hudson might not be admitted even after his next life. It would be against the rules, but she *will* overwrite his ethics module with a copy of her own. Her ethics module was obviously good enough to be admitted.

"We will be together—forever—next time," Victoria says with rays of hope and determination. She imagines hugging him even more tightly. ■

Tadayoshi Kohno is a professor in the Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA 98195 USA. His research interests include helping protect the security, privacy, and safety of users of current- and future-generation technologies. Kohno received a Ph.D. in computer science from the University of California, San Diego. He is a Fellow of IEEE. Contact him at yoshi@cs.washington.edu.