

Lecture 10.

Generative Agent-Based Models

CS 222: AI Agents and Simulations
Stanford University
Joon Sung Park

Announcements

AgentBank-CS222

- We will release to the class this early next week.
- If you want to [redact], please let us know.
- Please come to the office hour to go over your ideas with the teaching staff!

Announcements

We have *very* exciting guest lecturers!!



Meredith Ringel Morris

Director for Human-AI Interaction Research
Google DeepMind



Serina Chang

Assistant Professor
EECS, UC Berkeley

Today

Pt. 1. Generative Agent-Based Models (GABM)

Pt. 2. Ethics and considerations for *agent banks*

Today

Pt. 1. Generative Agent-Based Models (GABM)

Pt. 2. Ethics and considerations for *agent banks*

Last time, we talked about models of individuals

- Models of individuals predict the behavior of a particular person.
- This opens up genuinely new opportunities.

arXiv:2402.01662v2 [cs.CY] 8 May 2024

Generative Ghosts: Anticipating Benefits and Risks of AI Afterlives

Meredith Ringel Morris¹, Jed R. Brubaker²

¹Google DeepMind
²University of Colorado Boulder
merrie@google.com, jed.brubaker@colorado.edu

Abstract

As AI systems quickly improve in both breadth and depth of performance, they lend themselves to creating increasingly powerful and realistic agents, including the possibility of agents modeled on specific people. We anticipate that within our lifetimes it may become common practice for people to create a custom AI agent to interact with loved ones and/or the broader world after death. We call these *generative ghosts*, since such agents will be capable of generating novel content rather than merely parroting content produced by their creators while alive. In this paper, we discuss the design space of potential implementations of generative ghosts. We then discuss the practical and ethical implications of generative ghosts, including potential positive and negative impacts on individuals and society. Based on these considerations, we lay out a research agenda for the AI and HCI research communities to empower people to create and interact with AI afterlives in a safe and beneficial manner.

Introduction

The past few years have brought incredible growth in the capabilities of generative AI models, particularly large language models (LLMs) such as GPT-4 (OpenAI 2023a), Palm 2 (Anil et al. 2023), and Llama 2 (Touvron et al. 2023), though there has also been incredible progress in generative AI for the production of images (Ramesh et al. 2022), video (Singer et al. 2022), and audio (Borsos et al. 2023), as well as a new generation of multimodal models (Yang et al. 2023; Google DeepMind 2023) that combine functionality across several media categories. These models, in turn, have given rise to new types of *generative agents* (Park et al. 2023), simulacra that can produce believable human behaviors, including capabilities such as memory and planning. While still in their infancy, generative agents and related technologies are likely to increase in fidelity and popularity as underlying model capabilities improve and compute costs drop. For instance, in November 2023 OpenAI released GPTs (OpenAI 2023b), a no-code interface for people to develop agentic AIs.

As AI models increase the set of human capabilities they can faithfully reproduce (Morris et al. 2023; Bubeck et al. 2023), societal change is inevitable. For instance, experts anticipate that powerful AI systems may profoundly change disparate areas of society such as the labor market (Eloundou et al. 2023), the education system (Kasneci et al. 2023),

the pursuit of scientific knowledge (Morris 2023), and criminal activities (Ferrara 2023). In this paper, we discuss how advances in AI might change personal and cultural practices around death and dying.

We introduce the concept of *generative ghosts*, AI agents that represent a deceased person, and discuss why we anticipate such representations will become popular within our lifetimes. We explore the design space of possible instantiations of generative ghosts and consider both the benefits that might lead to their adoption and the practical and ethical concerns such technology may introduce.

Our contributions include: (1) identifying and characterizing an emerging phenomenon of creating “generative ghosts” to represent the deceased, (2) introducing a taxonomy of design dimensions and analysis of potential benefits and harms that can be used to support future empirical research and motivate fieldwork. By characterizing this emerging trend, highlighting potential risks to be averted, and creating a framework for future investigation, we aim to ensure that future technical and sociotechnical systems will maximize the potential benefits of “AI afterlives” while minimizing potential risks.

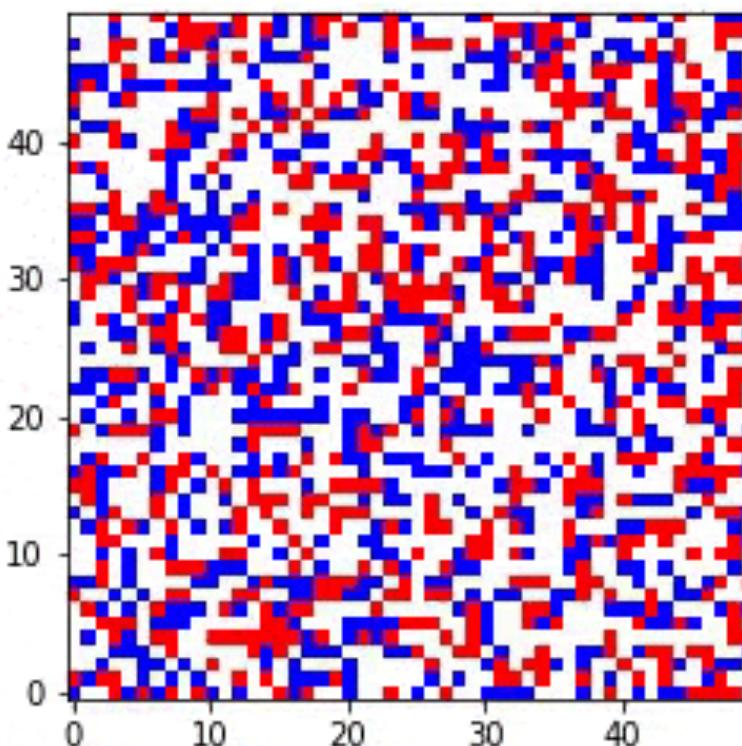
Related Work

We discuss the rich literature on how technologies have changed practices around death and dying and initial forays into AI afterlives by individuals and start-up ventures.

Post-Mortem Technology

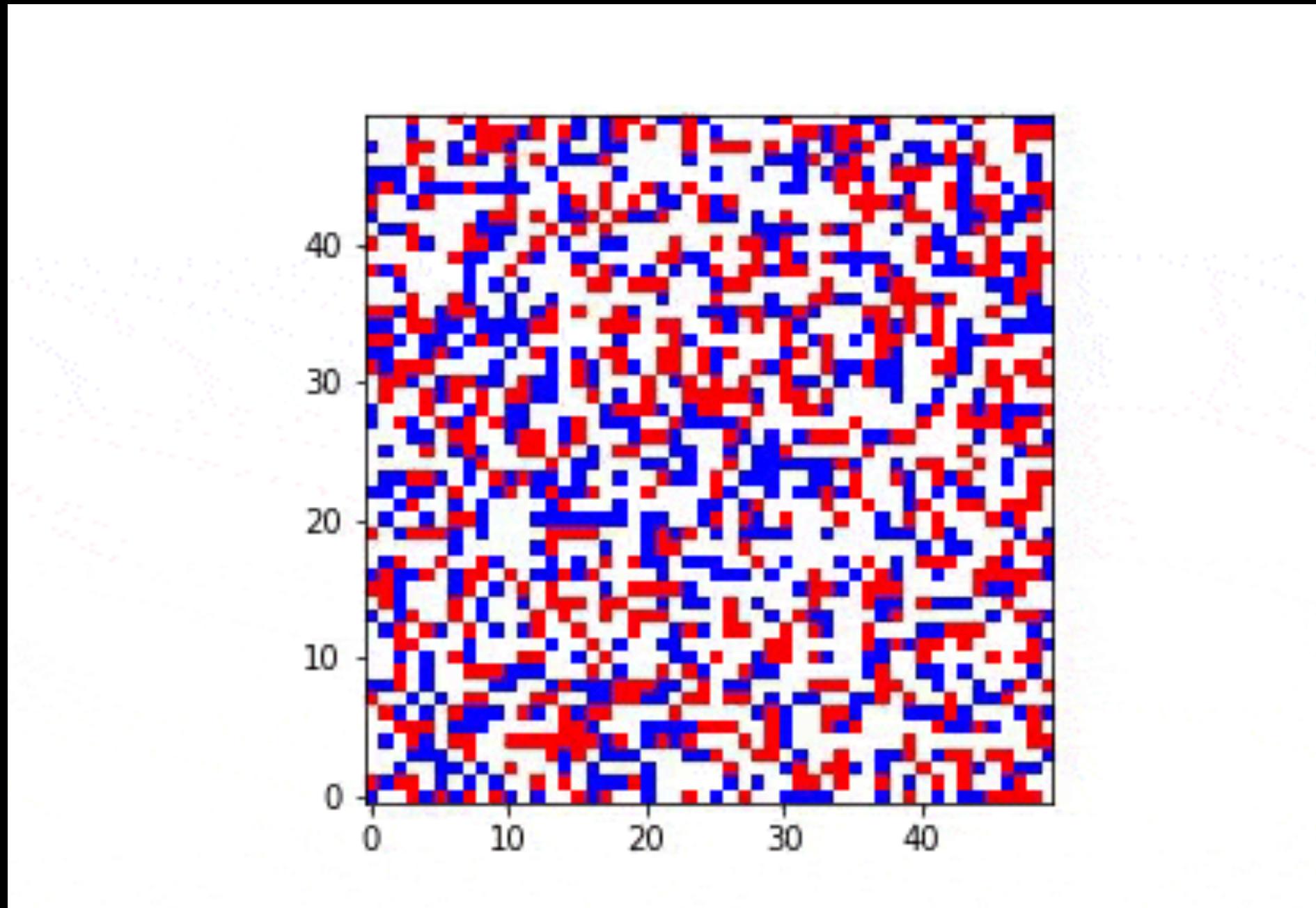
Throughout history, people have turned to technology to remember, memorialize, and even interact with the dead. Gravestones and other burial markers can be traced nearly back to 3000 B.C.E. (Taylor 2001). Obituaries in the U.S., while dating back to the 16th century, became more common during the 19th century in part due to the U.S. Civil War (Hume 2000) – an event that also brought embalming into favor. Even the medium of the Spiritualism movement in the late 19th and early 20th century turned to telegraphs, radio-wave detectors, and later wireless radio in their attempts to detect the presence of and communicate with the dead (National Science and Media Museum 2022).

During the earliest days of the World Wide Web, when people would create personal Home Pages describing their lives and family, it was routine for people to dedicate a page



T. C. Schelling, Dynamic models of segregation. J. Math. Sociol. 1, 143–186 (1971).

Agent-based models (ABM) studies the interaction between individual agents



T. C. Schelling, *Micromotives and Macrobbehavior* (W.W. Norton & Company, 1978).

T. C. Schelling, Dynamic models of segregation. *J. Math. Sociol.* 1, 143–186 (1971).

J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).

Q. What types of interactions do you see here?

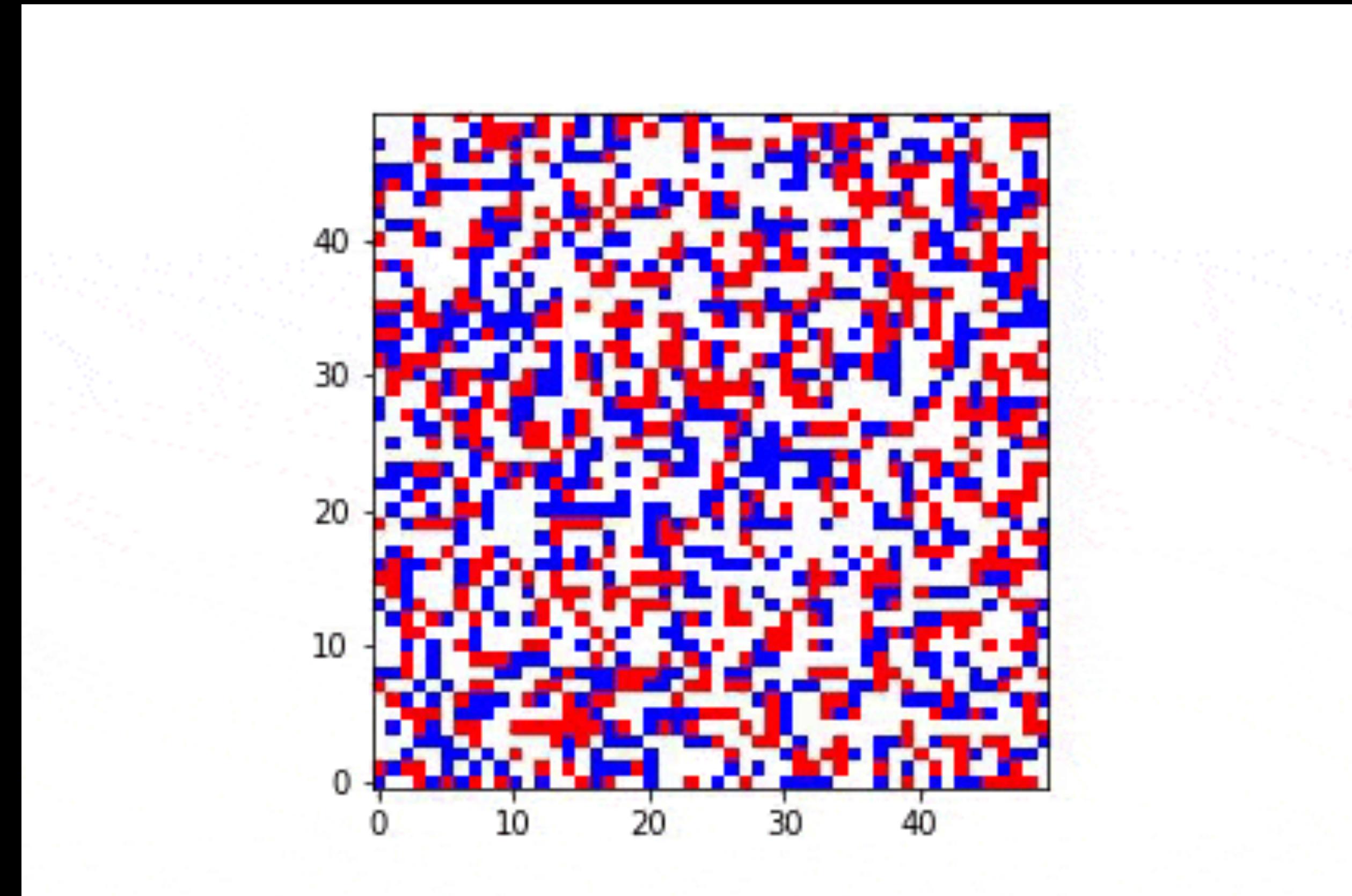
Class activity

Simulated voting (demo)

- You describe your “persona” that will be used to build your agent.
- We place your agent a room with other agents.
- The agents will decide who will be their leader.

Q. What kinds of findings might you see in GABM?

ABM often resulted in stable equilibria. What about GABM?



Model of segregation

Today

Pt. 1. Generative Agent-Based Models (GABM)

Pt. 2. Ethical considerations for *agent banks*

Genome sequencing: what was their contribution?

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

* A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

The rediscovery of Mendel's laws of heredity in the opening weeks of the 20th century^{1–3} sparked a scientific quest to understand the nature and content of genetic information that has propelled biology for the last hundred years. The scientific progress made falls naturally into four main phases, corresponding roughly to the four quarters of the century. The first established the cellular basis of heredity: the chromosomes. The second defined the molecular basis of heredity: the DNA double helix. The third unlocked the informational basis of heredity, with the discovery of the biological mechanism by which cells read the information contained in genes and with the invention of the recombinant DNA technologies of cloning and sequencing by which scientists can do the same.

The last quarter of a century has been marked by a relentless drive to decipher first genes and then entire genomes, spawning the field of genomics. The fruits of this work already include the genome sequences of 599 viruses and viroids, 205 naturally occurring plasmids, 185 organelles, 31 eubacteria, seven archaea, one fungus, two animals and one plant.

Here we report the results of a collaboration involving 20 groups from the United States, the United Kingdom, Japan, France, Germany and China to produce a draft sequence of the human genome. The draft genome sequence was generated from a physical map covering more than 96% of the euchromatic part of the human genome and, together with additional sequence in public databases, it covers about 94% of the human genome. The sequence was produced over a relatively short period, with coverage rising from about 10% to more than 90% over roughly fifteen months. The sequence data have been made available without restriction and updated daily throughout the project. The task ahead is to produce a finished sequence, by closing all gaps and resolving all ambiguities. Already about one billion bases are in final form and the task of bringing the vast majority of the sequence to this standard is now straightforward and should proceed rapidly.

The sequence of the human genome is of interest in several respects. It is the largest genome to be extensively sequenced so far, being 25 times as large as any previously sequenced genome and eight times as large as the sum of all such genomes. It is the first vertebrate genome to be extensively sequenced. And, uniquely, it is the genome of our own species.

Much work remains to be done to produce a complete finished sequence, but the vast trove of information that has become available through this collaborative effort allows a global perspective on the human genome. Although the details will change as the sequence is finished, many points are already clear.

- The genomic landscape shows marked variation in the distribution of a number of features, including genes, transposable elements, GC content, CpG islands and recombination rate. This gives us important clues about function. For example, the developmentally important HOX gene clusters are the most repeat-poor regions of the human genome, probably reflecting the very complex coordinate regulation of the genes in the clusters.
- There appear to be about 30,000–40,000 protein-coding genes in the human genome—only about twice as many as in worm or fly. However, the genes are more complex, with more alternative splicing generating a larger number of protein products.
- The full set of proteins (the 'proteome') encoded by the human genome is more complex than those of invertebrates. This is due in part to the presence of vertebrate-specific protein domains and motifs (an estimated 7% of the total), but more to the fact that vertebrates appear to have arranged pre-existing components into a richer collection of domain architectures.
- Hundreds of human genes appear likely to have resulted from horizontal transfer from bacteria at some point in the vertebrate lineage. Dozens of genes appear to have been derived from transposable elements.
- Although about half of the human genome derives from transposable elements, there has been a marked decline in the overall activity of such elements in the hominid lineage. DNA transposons appear to have become completely inactive and long-terminal repeat (LTR) retrotransposons may also have done so.
- The pericentromeric and subtelomeric regions of chromosomes are filled with large recent segmental duplications of sequence from elsewhere in the genome. Segmental duplication is much more frequent in humans than in yeast, fly or worm.
- Analysis of the organization of Alu elements explains the long-standing mystery of their surprising genomic distribution, and suggests that there may be strong selection in favour of preferential retention of Alu elements in GC-rich regions and that these 'selfish' elements may benefit their human hosts.
- The mutation rate is about twice as high in male as in female meiosis, showing that most mutation occurs in males.
- Cytogenetic analysis of the sequenced clones confirms suggestions that large GC-poor regions are strongly correlated with 'dark G-bands' in karyotypes.
- Recombination rates tend to be much higher in distal regions (around 20 megabases (Mb)) of chromosomes and on shorter chromosome arms in general, in a pattern that promotes the occurrence of at least one crossover per chromosome arm in each meiosis.
- More than 1.4 million single nucleotide polymorphisms (SNPs) in the human genome have been identified. This collection should allow the initiation of genome-wide linkage disequilibrium mapping of the genes in the human population.

In this paper, we start by presenting background information on the project and describing the generation, assembly and evaluation of the draft genome sequence. We then focus on an initial analysis of the sequence itself: the broad chromosomal landscape; the repeat elements and the rich palaeontological record of evolutionary and biological processes that they provide; the human genes and proteins and their differences and similarities with those of other

860 © 2001 Macmillan Magazines Ltd NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com

What are the parallels you see between the genome bank and the agent bank?

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

*A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

The rediscovery of Mendel's laws of heredity in the opening weeks of the 20th century^{1–3} sparked a scientific quest to understand the nature and content of genetic information that has propelled biology for the last hundred years. The scientific progress made falls naturally into four main phases, corresponding roughly to the four quarters of the century. The first established the cellular basis of heredity: the chromosomes. The second defined the molecular basis of heredity: the DNA double helix. The third unlocked the informational basis of heredity, with the discovery of the biological mechanism by which cells read the information contained in genes and with the invention of the recombinant DNA technologies of cloning and sequencing by which scientists can do the same.

The last quarter of a century has been marked by a relentless drive to decipher first genes and then entire genomes, spawning the field of genomics. The fruits of this work already include the genome sequences of 599 viruses and viroids, 205 naturally occurring plasmids, 185 organelles, 31 eubacteria, seven archaea, one fungus, two animals and one plant.

Here we report the results of a collaboration involving 20 groups from the United States, the United Kingdom, Japan, France, Germany and China to produce a draft sequence of the human genome. The draft genome sequence was generated from a physical map covering more than 96% of the euchromatic part of the human genome and, together with additional sequence in public databases, it covers about 94% of the human genome. The sequence was produced over a relatively short period, with coverage rising from about 10% to more than 90% over roughly fifteen months. The sequence data have been made available without restriction and updated daily throughout the project. The task ahead is to produce a finished sequence, by closing all gaps and resolving all ambiguities. Already about one billion bases are in final form and the task of bringing the vast majority of the sequence to this standard is now straightforward and should proceed rapidly.

The sequence of the human genome is of interest in several respects. It is the largest genome to be extensively sequenced so far, being 25 times as large as any previously sequenced genome and eight times as large as the sum of all such genomes. It is the first vertebrate genome to be extensively sequenced. And, uniquely, it is the genome of our own species.

Much work remains to be done to produce a complete finished sequence, but the vast trove of information that has become available through this collaborative effort allows a global perspective on the human genome. Although the details will change as the sequence is finished, many points are already clear.

• The genomic landscape shows marked variation in the distribution of a number of features, including genes, transposable elements, GC content, CpG islands and recombination rate. This gives us important clues about function. For example, the developmentally important HOX gene clusters are the most repeat-poor regions of the human genome, probably reflecting the very complex coordinate regulation of the genes in the clusters.

• There appear to be about 30,000–40,000 protein-coding genes in the human genome—only about twice as many as in worm or fly. However, the genes are more complex, with more alternative splicing generating a larger number of protein products.

• The full set of proteins (the 'proteome') encoded by the human genome is more complex than those of invertebrates. This is due in part to the presence of vertebrate-specific protein domains and motifs (an estimated 7% of the total), but more to the fact that vertebrates appear to have arranged pre-existing components into a richer collection of domain architectures.

• Hundreds of human genes appear likely to have resulted from horizontal transfer from bacteria at some point in the vertebrate lineage. Dozens of genes appear to have been derived from transposable elements.

• Although about half of the human genome derives from transposable elements, there has been a marked decline in the overall activity of such elements in the hominid lineage. DNA transposons appear to have become completely inactive and long-terminal repeat (LTR) retrotransposons may also have done so.

• The pericentromeric and subtelomeric regions of chromosomes are filled with large recent segmental duplications of sequence from elsewhere in the genome. Segmental duplication is much more frequent in humans than in yeast, fly or worm.

• Analysis of the organization of Alu elements explains the long-standing mystery of their surprising genomic distribution, and suggests that there may be strong selection in favour of preferential retention of Alu elements in GC-rich regions and that these 'selfish' elements may benefit their human hosts.

• The mutation rate is about twice as high in male as in female meiosis, showing that most mutation occurs in males.

• Cytogenetic analysis of the sequenced clones confirms suggestions that large GC-poor regions are strongly correlated with 'dark G-bands' in karyotypes.

• Recombination rates tend to be much higher in distal regions (around 20 megabases (Mb)) of chromosomes and on shorter chromosome arms in general, in a pattern that promotes the occurrence of at least one crossover per chromosome arm in each meiosis.

• More than 1.4 million single nucleotide polymorphisms (SNPs) in the human genome have been identified. This collection should allow the initiation of genome-wide linkage disequilibrium mapping of the genes in the human population.

In this paper, we start by presenting background information on the project and describing the generation, assembly and evaluation of the draft genome sequence. We then focus on an initial analysis of the sequence itself: the broad chromosomal landscape; the repeat elements and the rich palaeontological record of evolutionary and biological processes that they provide; the human genes and proteins and their differences and similarities with those of other

860 © 2001 Macmillan Magazines Ltd NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com



Who owns your agents?

LIVE — Medicare's two insurance systems: Current performance and options for reform

BROOKINGS Election '24 | U.S. Economy | International Affairs | Technology & Information | Race in Public Policy | Topics | Regions

RESEARCH Hollywood writers went on strike to protect their livelihoods from generative AI. Their remarkable victory matters for all workers.

Molly Kinder April 12, 2024

Growing up in a working-class immigrant family in Queens, N.Y., Danny Tolli dreamed of a career in television. "I was a storyteller straight out of the womb," Tolli told me in November. After studying film and television at New York University, he worked his way up in Hollywood, starting out as an assistant a decade ago and rising to a TV writer and co-executive producer today.

Until last year, Tolli—like most of us—never imagined that artificial intelligence might threaten his career. For decades, automation has concentrated in the kinds of routine, blue-collar jobs he grew up around. Highly creative and cognitive roles—like writing the stories that shape our culture—were

Workers and AI: Voices from the front lines of disruption

In this ongoing series, Brookings Metro introduces you to workers in occupations that will likely face disruption from generative AI, including writers, legal assistants, illustrators, accountants, and customer service representatives.

Generative Ghosts: Anticipating Benefits and Risks of AI Afterlives

Meredith Ringel Morris¹, Jed R. Brubaker²

¹Google DeepMind

²University of Colorado Boulder

merrie@google.com, jed.brubaker@colorado.edu

Abstract

As AI systems quickly improve in both breadth and depth of performance, they lend themselves to creating increasingly powerful and realistic agents, including the possibility of agents modeled on specific people. We anticipate that within our lifetimes it may become common practice for people to create a custom AI agent to interact with loved ones and/or the broader world after death. We call these *generative ghosts*, since such agents will be capable of generating novel content rather than merely parroting content produced by their creator while living. In this paper, we first discuss the design space of potential implementations of generative ghosts. We then discuss the practical and ethical implications of generative ghosts, including potential positive and negative impacts on individuals and society. Based on these considerations, we lay out a research agenda for the AI and HCI research communities to empower people to create and interact with AI afterlives in a safe and beneficial manner.

Introduction

The past few years have brought incredible growth in the capabilities of generative AI models, particularly large language models (LLMs) such as GPT-4 (OpenAI 2023a), Palm 2 (Anil et al. 2023), and Llama 2 (Touvron et al. 2023), though there has also been incredible progress in generative AI for the production of images (Ramesh et al. 2022), video (Singer et al. 2022), and audio (Borsos et al. 2023), as well as a new generation of multimodal models (Yang et al. 2023; Google DeepMind 2023) that combine functionality across several media categories. These models, in turn, have given rise to new types of *generative agents* (Park et al. 2023), simulacra that can produce believable human behaviors, including capabilities such as memory and planning. While still in their infancy, generative agents and related technologies are likely to increase in fidelity and popularity as underlying model capabilities improve and compute costs drop. For instance, in November 2023 OpenAI released GPTs (OpenAI 2023b), a no-code interface for people to develop agentic AIs.

As AI models increase the set of human capabilities they can faithfully reproduce (Morris et al. 2023; Bubeck et al. 2023), societal change is inevitable. For instance, experts anticipate that powerful AI systems may profoundly change disparate areas of society such as the labor market (Eloundou et al. 2023), the education system (Kasneci et al. 2023),

the pursuit of scientific knowledge (Morris 2023), and criminal activities (Ferrara 2023). In this paper, we discuss how advances in AI might change personal and cultural practices around death and dying.

We introduce the concept of *generative ghosts*, AI agents that represent a deceased person, and discuss why we anticipate such representations will become popular within our lifetimes. We explore the design space of possible instantiations of generative ghosts and consider both the benefits that might lead to their adoption and the practical and ethical concerns such technology may introduce.

Our contributions include: (1) identifying and characterizing an emerging phenomenon of creating "generative ghosts" to represent the deceased, (2) introducing a taxonomy of design dimensions and analysis of potential benefits and harms that can be used to support future empirical research and motivate fieldwork. By characterizing this emerging trend, highlighting potential risks to be averted, and creating a framework for future investigation, we aim to ensure that future technical and sociotechnical systems will maximize the potential benefits of "AI afterlives" while minimizing potential risks.

Related Work

We discuss the rich literature on how technologies have changed practices around death and dying and initial forays into AI afterlives by individuals and start-up ventures.

Post-Mortem Technology

Throughout history, people have turned to technology to remember, memorialize, and even interact with the dead. Gravestones and other burial markers can be traced nearly back to 3000 B.C.E. (Taylor 2001). Obituaries in the U.S., while dating back to the 16th century, became more common during the 19th century in part due to the U.S. Civil War (Hume 2000) – an event that also brought embalming into favor. Even the mediums of the Spiritualism movement in the late 19th and early 20th century turned to telegraphs, radio-wave detectors, and later wireless radio in their attempts to detect the presence of and communicate with the dead (National Science and Media Museum 2022).

During the earliest days of the World Wide Web, when people would create personal Home Pages describing their lives and family, it was routine for people to dedicate a page

Kinder, M. (2024). Hollywood writers went on strike to protect their livelihoods from generative AI. Their remarkable victory matters for all workers. Brookings. April 12.
M. R. Morris, J. R. Brubaker, Generative Ghosts: Anticipating Benefits and Risks of AI Aerlives. Preprint (2024).

arXiv:2402.01662v2 [cs.CY] 8 May 2024

References

- T. C. Schelling, Dynamic models of segregation. *J. Math. Sociol.* 1, 143–186 (1971).
- T. C. Schelling, *Micromotives and Macrobbehavior* (W.W. Norton & Company, 1978).
- J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, M. S. Bernstein, Generative agents: Interactive simulacra of human behavior, in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (ACM, 2023).
- International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921. <https://doi.org/10.1038/35057062>
- Kinder, M. (2024). Hollywood writers went on strike to protect their livelihoods from generative AI. Their remarkable victory matters for all workers. Brookings. April 12.
- M. R. Morris, J. R. Brubaker, Generative Ghosts: Anticipating Benefits and Risks of AI Aerlives. Preprint (2024).

CS 222: AI Agents and Simulations

Stanford University

Joon Sung Park