NATIONAL ENDOWMENT FOR THE HUMANITIES

DIVISION OF RESEARCH PROGRAMS

Narrative Section of a Successful Application

The attached document contains the grant narrative of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and applicants are urged to prepare a proposal that reflects their unique project and aspirations. Prospective applicants should consult the Research Programs application guidelines at https://www.neh.gov/grants/research/collaborative-research-grants for instructions. Applicants are also strongly encouraged to consult with the NEH Division of Research Programs staff well before a grant deadline.

The attachment only contains the grant narrative, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

The application format has been changed since this application was submitted. You must follow the guidelines in the currently posted Notice of Funding Opportunity (see the Notice posted on the Collaborative Research program page linked above).

Project Title: Unboxing Artificial Intelligence: An International Collaboration Bringing Humanities Perspectives to AI

Institution: Rutgers University

Project Director: Lauren Goodlad

Grant Program: Collaborative Research

Project Narrative - Unboxing AI: A New International Scholarly Collaboration between Rutgers University, New Brunswick and the Australian National University

Significance and Impact: Artificial Intelligence (AI) is an emergent set of computer technologies that affect individuals, communities, and societies at a global scale. Touted as a fourth industrial revolution, AI is, nonetheless, poorly understood and subject to hype, misinformation, and anxiety. Though there is increasing talk about making AI "ethical," "democratic," and "human-centered," scholars in the humanities seldom shape these discussions. In the words of ethics professor Amy Webb, no one has yet been "incentivized to consider the unforeseen costs of optimizing AI in the absence of codified, humanistic principles." The result is conversations about AI that lack fundamental understanding of this profoundly social phenomenon.

With growing public awareness of AI's role in surveillance, bias, political polarization, and environmental harm, discussions over how to govern AI have begun to permeate research. Critics have projected the idea of AI as an opaque "Black Box." Though useful to a point, this metaphor suggests a problem of algorithmic impenetrability to be resolved through technical fixes that offer "explainable" software and "transparent" data processing. Our international collaboration seeks to "unbox" AI more holistically. AI, we contend, is not reducible to an opaque algorithm in a metaphorical box; and questions over how data are collected, curated, labeled, and computed are much larger than any single digital process. Rather, AI involves humans, machines, and their shared environments in complex interrelations of public and private, local and global, nature and culture, technology and power, metrics and code. What is needed more than ever is an interdisciplinary and collaborative approach to initiate dialogue, articulate "humanistic principles," and make AI a topic of wide-ranging study in conversation with many stakeholders--endeavors our research questions are designed to spark.

"Unboxing AI" is a new scholarly collaboration between faculty at two major universities and their networks. At Rutgers, Project Director Lauren M. E. Goodlad, is leading an interdisciplinary working group in Critical AI that brings together the Center for Cultural Analysis (CCA) and the Center for Cognitive Science (RuCCS) to foster public events, research, and teaching opportunities. At Australian National University in Canberra, Lead International Collaborator Katherine Bode, under the auspices of a four-year Future Fellowship, is working with faculty in ANU's Computational Culture Lab in dialogue with two other leading ANU organizations: the Humanising Machine Intelligence network and the 3Ai Institute.

From October 2021 through August 2022, we will establish the collaboration and plan its activities. Our work will center on two exploratory workshops that delve into research questions in conversation with leading AI thinkers. Several meetings will be live-streamed and archived; student participants will be invited to blog and tweet. Our written plan (due August 2022) will lay out steps for a peer-reviewed special issue (including abstract and table of contents) and may also project 1) a jointly-hosted international conference, 2) jointly-developed and/or team-taught curricula for ANU/RU students, and 3) a white paper for international circulation.

Substance and Context: Like most technologies, AI has the potential to be democratic, inclusive, fair, and environmentally sound. But its leading purveyors are for-profit enterprises with no incentive to share the proprietary algorithms or curb the profitable surveillance that fuels the technology as it currently exists. The metaphor of AI as a "Black Box" inadvertently favors this status quo by perpetuating the idea that "explainable" software will readily pave the way for "ethical" outcomes.

Misleading information about AI's achievements circulate widely through press releases, social media, journalism, and even some peer-reviewed publications. To complicate matters, a confusing specialist rhetoric of "neural" networks and "deep learning" leads many to assume that AI's software architectures replicate the human brain. In actuality, the current technology excels at artificial *narrow* intelligence (ANI) but does not advance a (human-like) artificial *general* intelligence (AGI) in any appreciable form. Impressively adept at particular tasks (such as detecting some cancers, translating text, or playing Go), ANI works by processing huge troves of data at unprecedented speed. Lacking sentience, emotion, common sense, causal reasoning, imagination, and a concept of the world, these powerful pattern-finders and predictors have no capacity to cognize the data points they plot and model. Defeating the world's best Go player, they lack a sense of "winning" or "play"; translating French into English, they do not "understand" either language. As computer scientist and philosopher Judea Pearl emphasizes, ANI cannot ask "why?" or "what if?" Still less can it adjudicate the social or moral consequences of its findings or activities.

Though some researchers predict that ever larger datasets and more powerful computers will eventually usher in AGI, Pearl is one of many to disagree. To be capable of human-like reasoning, he argues, computers need more than data-mining: they also need models of real-world knowledge and experience. Given that understanding of the brain is itself incomplete, technologists are not well-positioned to "reverse engineer" its manifold functions (Schneider). Nonetheless, major figures like the scientist Stuart Russell and the philosopher Nick Bostrom look past this impasse: although technology has yet to produce AGI, they worry about an artificial *superintelligence* that may one day outstrip human intelligence (Bostrom, Russell, Kurzweil). As we see it, however, the most imminent concern is not that machine intelligence will outstrip human capabilities (which in narrow applications it already does), but, rather, that human imagination, causal inference, and counterfactual reasoning is being depreciated by a culture that overemphasizes the data-driven "intelligence" at which machines excel. At best, writes mathematician Cathy O'Neil, data-centric technologies "codify the past." To ensure a future of democratic flourishing requires a "moral imagination" that data cannot provide.

At the same time, the hype over what AI may do in the distant future, whether utopian or apocalyptic, inhibits robust understanding and discussion in the here-and-now. Consider HireVue, a leading job interview application that has already assessed the videos of more than a

million job-seekers. Sold as an objective tool for measuring "employability" through facial movements, tone of voice, and mannerisms, the company's tech, according to computer scientist Meredith Whittaker, is a pseudoscientific "license to discriminate" (Harwell). Already contending with news of an AI-powered "age of surveillance capitalism" (Zuboff), biased data-sets (O'Neil, Noble), and the polarizing effects of social media algorithms (Vaidhiyanathan), the general public has yet to grasp the myriad challenges that AI presents. The Silicon Valley model of AI's advent positions citizens as consumers of technology, not participants in a democratic process. Thus, plans for self-driving cars evolve with little or no input from environmentalists, urban planners, or ordinary people. Instead, tech companies look to the public to subsidize necessary infrastructure and to help them identify the small fixes needed to make their products more "explainable" and "human-centered."

While humanists can shine powerful light on this dilemma, their activities so far have been limited. To be sure, scientists are increasingly encouraged to add humanists to their teams; and corporations such as Microsoft offer fellowships to scholars of digital culture. But such roles position humanists as support players in a narrative that upholds innovation, disruption, and private profit as the main drivers of change. At Stanford's Institute for Human-Centered Artificial Intelligence (HAI), which promises AI research "to improve the human condition," co-director Fei-Fei Li, a roboticist, collaborates with several humanists. Though HAI strives to bring women and minorities into tech, Li's rather minimalist notion of "democratizing AI," developed when she was a chief scientist for Google, entails "creating products that empower businesses and partners" and "taking feedback from customers" (qtd. in Ford).

Even Wai Chee Dimock's well-intentioned *PMLA* editor's column inadvertently reproduces hype: it reports that AI "might render educated human beings superfluous"--an astounding claim--and introduces Elon Musk ("one who know[s] the technology from the inside") as if he were a neutral authority on such prognoses. Building on a Microsoft presentation during the January 2020 Modern Language Association convention, Dimock cites an op-ed in which the founder of a start-up promises that, within five years, AIs will be writing screenplays deemed "better than human writing" (Lea). In actuality, the text-generating software in question, OpenAI's GPT-2 and 3, is nowhere near delivering on that expectation. Rather, despite millions of dollars, "breathtaking amounts of carbon emissions," and 450 gigabytes of data, GPT-3, according to cognitive psychologist Gary Marcus, is a "fluent spouter" of statistically probable language, not a "reliable interpreter of the world" (Marcus "GPT-3"; *Rebooting*). Still less is it a source for fabulous screenplays. Like the promise that safe driverless cars are just around the corner, ready to replace public transit with fleets managed by Uber, Tesla, or Google's Waymo, such hype encourages technological determinism and civic passivity. It ensures that as soon as a new product is ready to "upend" existing markets and ways of life, investors will profit while taxpayers absorb the social and environmental costs.

As an interdisciplinary humanities collaboration, Unboxing AI is committed to rigorous open-mindedness. Far from anti-technology, our exploratory workshops will put leading technologists alongside humanist counterparts in fields such as anthropology, history, law, and philosophy. Nonetheless, as against the culture of hype and the subordination of humanists to support roles, our research questions put critical humanities thinking at the fore. Our diverse readings (see Attachment 5) range from Alan Turing's classic 1950 essay on the "imitation game," and anthropologist Kate Crawford's "Anatomy of an AI System," to legal scholar Rashida Richardson's "Confronting Black Boxes" (on the racial and civil rights implications of data-driven AI). Such wide-ranging reading--culled to evoke a field of study that is just beginning to coalesce--will inform two exploratory workshops built around visits from leading thinkers and centered on the following core research questions. What, we will ask, can interdisciplinary dialogue do to shape and articulate the humanistic principles on which a democratic AI can be grounded and developed? Given that explaining the "Black Box" of software architectures proffers a merely technical fix for a broadly social phenomenon, what kinds of collaborative research, pedagogy, public engagement, and policy will help to broaden conversations, enhance equitable goals, and open AI research and implementation to the democratic process? Might humanists help AI researchers to invent a different kind of "reverse engineering"--one that begins with models for social justice, environmental sustainability, and intellectual flourishing and designs from there?

Methods and Execution: When research on AI makes a claim to interdisciplinarity, that often means that a group of technologists has added a token philosopher, anthropologist, or literary critic to their team. In this way, humanist input is thought to provide a handy set of guardrails that temper technological development and assure "ethical," or "human-centered" outcomes. This additive approach fails to acknowledge that many of AI's core issues fall squarely within a humanities tradition: from the preparation of textual corpora for algorithmic "deep learning," the historicity of language and meaning, and the embedment of racism or misogyny in data, to the invariably social and cultural dimensions of knowledge production. Unboxing AI begins with the recognition that no superficial engagement of the humanities will meet the challenges that AI presents. Our goal is to plan and establish a scholarly collaboration built on robust interdisciplinary methods**:** we will put scholars from across the humanities into extended conversations with computer scientists, mathematicians, data scientists, computational linguists and others.

Both exploratory workshops will ensure interdisciplinarity through diverse reading and invited presentations, discussed with the members of the collaborative and available to others through live-streaming, blogs, social media, and a digital archive. Our workshop format is modeled partly on the approach of journals such as *New Literary History*, where special issue editors invite select scholars to discuss and refine ideas. This format will not only provide stimulating encounters for participants, but also ferment the "Unboxing AI" special issue we project for the

next stage of collaboration. By exploring core histories and research questions across fields and disciplines, we hope to break new ground--providing generative models for turning this so-called fourth industrial revolution into fertile terrain for humanistic engagement. Instead of organizing our meetings by discipline, we will work through overlapping focal points that cut across disciplinary silos and dig into common research concerns: *algorithms*, *data curation*, *histories, law, media and design*, *ecosystems and complexity*, *race and justice*, and *structures and society* (see also Attachment 6).

Our open-ended discussions will encourage participants to respond to earlier presentations while testing out new ideas and pathways throughout. The goal is to identify and articulate the humanistic foundations necessary to understanding AI (as technology, commercial product, social practice, research tool, and cultural phenomenon) in order to help shape its development. To do so, we will study the historical and material conditions of AI as it now exists and as it might be, while also considering how AI, in turn, may impact the humanities now and in the future. Rather than seeing these virtual meetings as makeshift substitutes for the kind of real-world engagements we hope to plan in the years ahead, our approach will take full advantage of the affordances of digital technologies: we will make the cultivation of ideas across a series of "virtual" encounters, part of the knowledge we produce. Beyond projecting the publication of a special issue, we hope to empower a new generation of interdisciplinary humanists who are conversant with AI and its challenges, as well as to enable technologists to integrate humanities insights into their research.

Although Unboxing AI is too interdisciplinary to single out any one theoretical framework, a consistent focus is the critical engagement of *ontology*—a term that computer scientists use to denote encoded representations of knowledge and models of the material world. While humanists often perceive *ontology* as an exclusively metaphysical concern, we invoke a *critical ontology* to emphasize that the reality that AI strives to grasp is, in the words of political theorist Alex Callinicos, "complex, structured, and multi-levelled"; or in Georg Lukács's formulation, a "configuration of complexes within a historical process." Such an approach recognizes that, just as objects of knowledge are embedded in material conditions, so the knowledge-making practices that scholars mobilize to describe, theorize, and model these objects are likewise embedded. Thus, rather than statically "applying" theories *to* AI, or "applying" AI *to* objects in the world, our critical framework acknowledges that all knowledge-making practices contribute to forming the phenomena they study. That is true whether the practice in question involves actor-network theory, historical materialism, or the "object-oriented" structure of programming languages such as Java.

From week to week, our discussions will traverse disciplinary perspectives and scholarly domains in the effort to bring humanistic insights and imperatives to the technological vision of a "human-centered" AI. We will learn about the gift economies of data-gathering (sociologists

Marion Fourcade and Daniel Kluttz); the granular "submarine" ontologies that visual AI opens to human study (philosopher and cognitive scientist Brian Cantwell Smith); the built-in exclusions of the historical archives adopted for computational analysis (digital humanist Lauren Klein); the limitations of data-centric approaches to knowledge (artist James Bridle, data scientist Cathy O'Neil, and computer scientist Judea Pearl); the exploited human labor that invisibly supports supposedly autonomous technologies (anthropologist Mary Gray and computer scientist Siddharth Suri); and the inherited racial structures that ostensibly neutral technologies reproduce despite ongoing efforts to "explain" "Black Boxes" and mathematically rectify biased data (applied mathematician and philosopher Lily Hu, information scientist Safiya Noble, computer scientist Rediet Abebe and economist Maximilian Kasy).

Rutgers and Australian National University (ANU) are ideal institutional partners for this international collaboration. The United States and Australia are major global players in AI scholarship. Rutgers is an emerging hub for interdisciplinary AI research, hosting a community of intellectuals across the humanities, arts, and sciences brought together through a recent partnership between the Center for Cultural Analysis (CCA) and the Rutgers Center for Cognitive Science (RuCCS). Through a steering committee chaired by Lauren Goodlad, the Critical AI working group has begun to organize events and to build an externally-facing digital platform for shared scholarship and dialogue. CCA has a longstanding record of fostering such interdisciplinarity including strong affiliations to RU's highly regarded programs in English, History, and Philosophy. RuCCS's cutting-edge work on the frontiers of cognitive science includes the recent hosting of the 7th International Conference on Motion and Computing. Critical AI's working group also includes partners in the School of Media and Information, the Blaustein School of Planning and Public Policy (which houses the Rutgers Center for Green Building), the Mason Gross School of the Arts (which houses the Art & Artificial Intelligence Lab), as well as faculty in Computer Science and the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS). The Australian National University has invested millions of dollars in AI research, hosting three centers, each based in one of the three major faculty areas: the Centre for Computational Culture, for the humanities and arts; the Humanising Machine Intelligence group, for the social sciences; and the 3Ai institute, for the sciences. Katherine Bode, whose research group pursues machine learning as a method for literary research, has collaborative relationships with members of all these centers; her hope is that Unboxing AI will enable these ANU partners to focus their activities on the role of humanities principles and frameworks in AI research.

These complementary research agendas make Rutgers and ANU ideal partners to progress debate, address core questions, and forge a new and versatile collaboration that synergizes the strengths of these two institutions.

History of the Project and Its Productivity: The idea for a Rutgers-ANU collaboration began when Lauren Goodlad, invited to speak at the Australian Literary Association's 2018 meeting, hosted by ANU, took the opportunity to explore common research interests with Katherine Bode. Goodlad was already an admirer of Bode's work on the importance of curating literary archives for computational analysis. As part of her ongoing efforts to assemble the Critical AI working group at Rutgers, Goodlad worked with CCA's director, Colin Jager, to invite Bode for a January 2020 event: "DH Futures: a Conversation about Archives, Data, and the Digital Humanities" (RU digital librarian Francesca Giannetti, RU literary scholar Andrew Goldstone, and DH scholar Yohei Igarashi from the University of Connecticut also took part). At about the same time, Bode and Goodlad spoke at MLA, enabling them to discuss the AI-related panel alluded to in Wai Chee Dimock's *PMLA* column (discussed above). Both were struck by the misleading claims about human-level AI which some participants conveyed.

Already exchanging work and learning more about their respective intellectual networks, Bode and Goodlad recognized that NEH's program for planning a new international collaboration could help them to formalize and broaden their evolving AI-related projects. They chose "Unboxing AI" to provide a galvanizing theme for two exploratory workshops leading up to a special issue through a collaboration that might eventually include an international conference, a series of experimental courses, and/or a white paper. The idea of "unboxing" AI acknowledges the problem of algorithmic opacity while simultaneously pointing to more complex social challenges that humanities scholars are well-positioned to elaborate and engage.

At ANU, Bode has a longstanding collaboration with researchers from the Computational Culture Lab, a collective of digital designers and artists who use AI in their critical and creative works, including co-director Baden Pailthorpe and core members Geoff Hinchcliffe and Mitchell Whitelaw. Bode is also in contact with several AI researchers at ANU, whom she anticipates will take part in the exploratory workshops: these include Tom Gedeon, Professor of Computer Science and Head of the Human Centred Computing (HCC) Research Area; Adrian McKenzie, Professor of Sociology and author of *Machine Learners: Archaeology of a Data Practice* (MIT Press 2010); and Seth Lazar, Professor of Philosophy and Director of Humanising Machine Intelligence (HMI).

At Rutgers, Goodlad has been working closely with Sara Pixley (Executive Director of RuCCS and a researcher in cognitive psychology and neuroscience); Atif Akin (whose work on AI combines technology, engineering and art); Brian McLaughlin (former faculty director of RuCCS, philosopher of mind, and expert in the ethical use of robots); Alex Guerrero (a moral philosopher and epistemologist); Matthew Stone (Chair of Computer Science and specialist in the effort to design "meaningful" AI); David Pennock (computer scientist and Director of DIMACS, whose own AI research integrates social science perspectives); and Colin Jager (a literary scholar who directs CCA). Other members of the Critical AI steering committee whom

Goodlad anticipates will take part in one or both workshops are Clint Andrews (an engineer, urban planner, and environmental policy-maker); Elisabeth Camp (philosopher of language and mind); Omar Dewachi (medical anthropologist and specialist in biopolitics); John McGann (neuroscientist and current faculty director of RuCCS); and Jamie Pietruska (historian of technoscience). Goodlad believes that Critical AI's virtual event series in February and March 2021, as well as the digital platform now being created to support it, will provide useful networks and infrastructures that should enable the new Unboxing AI collaboration to focus on studying research questions, organizing workshops, generating material for a groundbreaking special issue, and envisioning next steps.

Bode and Goodlad project that they will serve as joint editors for the "Unboxing AI" special issue. A key milestone will be the preparation of content for peer review no more than six months after completion of the written plan in August 2022. Bode and Goodlad will eventually seek external funding for release time, travel, attendance of major conferences (including those for AI technologists), the organizing of an international symposium, joint pedagogical experiments, and a workshop for white-paper writing. With the running start afforded by Rutgers' and ANU's strengths in this emerging area, and the boost of an NEH-funded international collaboration, Bode and Goodlad anticipate that Unboxing AI will continue to develop high-impact activities and grant-seeking opportunities through AY 2025-6.

Collaboration: Lead Project Director Lauren Goodlad, a professor of English and Comparative Literature and a faculty affiliate of CCA, RuCCS, and the Center for British Studies (RBSC), has a proven history of working collaboratively on interdisciplinary projects. At the University of Illinois, Urbana she directed the Unit for Criticism & Interpretive Theory, an internationally renowned center for research in the humanities and interpretive social sciences, between 2008 and 2014. Under Goodlad's direction, the Unit for Criticism published a highly-regarded essay collection with Duke University Press; multiple special issues (in journals including the *Journal of Human Rights, Novel: A Forum on Fiction,* and *Victorian Studies*); applied for and oversaw internal funding for faculty research fellowships; and won and oversaw a Mellon Foundation programming grant. She has worked closely with anthropologists, art historians, critical race theorists, historians, media scholars, political theorists and, most recently, with computer scientists, DH scholars, philosophers, and linguists. As Provost Fellow for Undergraduate Education between 2014 and 2016, Goodlad also led a new program in Grand Challenge Learning which entailed campus-wide general education clusters for undergrads on the topics of Health & Wellness, Inequality, and Sustainability. She is currently chair of the MLA's forum on Literature and History, serves on the boards of several journals, and is the co-editor of a forthcoming December 2020 special issue of *MLQ* titled *What Is and Isn't Changing: Critique After Postcritique.* Since moving to Rutgers in 2017, Goodlad has served as associate chair of English and a member of the department's executive committee for graduate studies. In addition to chairing the Critical AI steering committee, she is a member of the executive committees for CCA and RBSC. Under the auspices of the latter, Goodlad worked with faculty in History to

organize a November 2020 event, “The Black Atlantic in the Age of Black Lives Matter.” As Lead Project Director for Unboxing AI, she will work closely with Katherine Bode to assure that decision-making for the new collaboration is conducted in consultation with all participants and stakeholders. Goodlad’s frequent teaching on literature and artificial intelligence enables her to bring highlights from her research into the classroom and vice versa. She anticipates working part-time on Unboxing AI from October 2021 through May 2022 and full-time thereafter in order to bring the August 2022 written plan to timely fruition.

Lead International Collaborator, Katherine Bode, a professor of Literary and Textual Studies, has an extensive history of productive, cross-disciplinary collaboration. From 2011 to 2013 she directed the Centre for Digital Humanities Research, an interdisciplinary group of researchers focused on expanding humanities approaches to digital technologies and digital approaches to humanities phenomena. In that position, she led major national and international projects including as Principal Investigator on the Mellon-funded Project Bamboo and as a Lead Investigator on Seeding the Commons (a collaboration with the Australian National Data Service to establish infrastructure and protocols for humanities data). Bode has been highly successful gaining funding from the Australian Research Council to establish and lead interdisciplinary research collaborations, including *To be continued* (which used machine learning methods to identify and explore over 20,000 literary works in digitized historical Australian newspapers) and *Reading at the Interface* (a project devoted to curating and theorizing vast new digital archives of literary reception). *To be continued* has received additional funding to continue the work of a bibliographer and two developers until 2024 in collaboration with the National Library of Australia. Beyond building and leading interdisciplinary research teams, Bode has co-convened several major conferences, including the first to combine Australian literary studies and eResearch (in 2009) and the inaugural event of the Australasian Association for Digital Humanities (in 2012), a professional association she helped to establish. Her research has been highly interdisciplinary: for instance, an essay co-authored with Tara Murphy, Professor of Astrophysics, and research presentations together with education researchers (such as Professor Wayne Sawyer) and librarians (such as Deputy Director of the National Library of Australia, Alison Dellit). Bode’s collaborative capabilities are further demonstrated by the co-editing of five collections, including *Resourceful Reading: The New Empiricism, eResearch and Australian Literary Culture* (2009), *Advancing Digital Humanities Research: Research, Methods, Theories* (2014) and *To be continued: The Australian Newspaper Fiction Database* (2018). She is Series Editor of Cambridge University Press’s Elements in Digital Literary Studies and on the editorial board of journals and organizations including *Post45 Data Collective*, *Humanist*, *International Journal of Digital Humanities*, and *Journal of Cultural Analytics*. Recently elected as a fellow of the Australian Academy of Humanities, Bode is also an Australian Research Council Future Fellow (2018 to 2022), on that organization’s College of Experts, and a visiting fellow at King’s College London from 2019 to 2021. Bode anticipates working part-time on Unboxing AI from October 2021 through May 2022 and full-time in August 2022 to complete the project’s written plan in a timely fashion.

In their respective roles as Lead Project Director and Lead International Collaborator, Goodlad and Bode will co-direct Unboxing AI. Rutgers graduate student Jacob Romanow is projected to serve as Project Assistant (performing a range of administrative tasks, including organizing meetings, maintaining the project website, uploading readings, organizing grad student blogging, and helping to publicize discussions and presentations). Bode and Goodlad will begin their work by establishing an interdisciplinary Advisory Board composed of key collaborators at Rutgers and ANU as well some leading international figures (see also Attachment 1). The Advisory Board will provide guidance and feedback on all planning, progress, and follow-up activities up to and including completion of the written plan in August 2022.

Work Plan: As fully detailed in Attachment 3, our work plan extends from October 1, 2021 through the end of August 2022. We will begin by establishing the collaboration, and create an Advisory Board composed equally of Rutgers and ANU members (with some leading AI figures from other institutions). Throughout October 2021, the project team (including the grad student Project Assistant) will work with the board and other key collaborators to organize the two exploratory workshops and prepare the final written plan. We will hold three planning meetings that include the project team, board, and other key collaborators so that all stakeholders have ample time to deliberate and plan before each of the two workshops and to take stock and make adjustments as needed.

Work plan tasks include setting up the website, inviting presenters, archiving readings, arranging for expert moderation and facilitation for each of the workshop meetings, publicizing events, organizing student blogging, archiving recorded presentations, and making decisions about the content of and venue for the projected special issue.

The two exploratory workshops will each feature eight meetings devoted to four different focal points. Four of the meetings for each workshop will be devoted to discussing readings, alternating with four additional meetings devoted to invited presentations (thus, each focal point will feature a discussion meeting followed by an invited presentation meeting). The 8 weekly meetings that comprise Workshop I will be scheduled during November/December 2021; the 8 additional weekly meetings that comprise Workshop II will take place during March/April 2022.

A final meeting will be scheduled in May 2022 for the purpose of collaborating on the drafting of the written plan: at this point, the board and other key partners (including graduate student members) will help the project team to decide on future steps for the collaborative above and beyond the special issue (e.g., international conference, grant-seeking, joint pedagogy, white paper). The written plan will include a complete abstract and table of contents for the special issue.

Final product and dissemination: Unboxing AI is committed to publicizing our workshops in advance so as to include as many interested researchers as possible in our live-streams and online discussions. Our extensive digital archive, blog series, and social media posts will continue the work of dissemination asynchronously. In addition, Bode, Goodlad, and select collaborators from among the group plan to attend relevant conferences in both Australia and the United States, especially those frequented by AI technologists.

Faculty at both ANU and RU, many of whom are already in dialogue with Bode and Goodlad about the new Unboxing AI collaboration, are engaged in relevant research projects (see also Attachment 1). This makes our plans to establish a new scholarly collaboration certain to succeed in forming two stimulating and well-publicized workshops, a generative and accessible digital archive, and a productive written plan with clearly laid out next steps.

Our intention to follow up immediately with an interdisciplinary special issue of a peer-reviewed journal, to be titled "Unboxing AI," is, we believe, highly feasible given that both Bode and Goodlad are experienced editors who have worked successfully with numerous journals and presses. In addition, Goodlad and Colin Jager are working with Duke University Press on a new journal devoted to AI-related subjects, with Bode already agreeing to serve as Associate Editor for Digital Humanities--one among several possible venues for a peer-reviewed "Unboxing AI" special issue. Beyond the latter publication, potential next steps could include a jointly-organized conference, a program of experimental courses for RU and ANU students, and/or a white paper. Bode and Goodlad will discuss these possibilities with their Advisory Board and other Unboxing AI participants during workshops and dedicated meetings in preparation for elaborating these projects more fully in the August 2022 written plan.

Potential grant opportunities for Unboxing AI after August 2022 might include a collaborative conference grant from NEH and/or grant-seeking for other activities from the Australian Research Council, the Mellon Foundation, and/or private foundations that offer dedicated funding for high-impact research in technology and AI.

Building on the momentum of the two workshops, digital archive, written plan, and the projected special issue and other potential follow-up activities, Bode and Goodlad conceive Unboxing AI as a lasting international collaboration that, we hope, will influence humanists and technologists on our respective campuses as well as beyond them.