Envisioning a Cancer AI Accelerator

Meeting Summary

January 10, 17, & 26

2023

Lee Cooper, Northwestern University
Jennifer Couch, NCI
Jun Deng, Yale University
Chris Gibbons, MD Anderson
Emily Greenspan, NCI
Sean Hanlon, NCI
Roxanne Jensen, NCI
Juli Klemm, NCI
Bill Lotter, Harvard
Anant Madabhushi, Emory University
Gina-Maria Pomann, Duke University
Catherine Schweppe, NCI

Table of Contents

I. Preface

In response to the community identified opportunity to build trans-disciplinary collaborations and accelerate the impactful and appropriate use of Artificial Intelligence (AI) in cancer research, the NCI is exploring the idea of implementing a "Cancer AI Accelerator." The Accelerator is envisioned as a nimble and adaptive organization providing connection and collaboration among researchers in the rapidly moving field of AI applied to cancer research and clinical practice as well as to fostering interactions with relevant organizations and initiatives beyond NCI. It is also NCI's intention that the Accelerator goals and activities align with several recent White House and federal government policies and guidelines related to AI, including the NIST Artificial Intelligence Risk Management Framework, Executive Order 13960 Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, the AI Bill of Rights, and the National Artificial Intelligence Research and Development Strategic Plan 2023 Update.

In January 2023, 35 academic thought-leaders in cancer-focused AI research were invited to participate in a series of three virtual working meetings to explore this opportunity and to envision the core structure, scope, and priorities for a Cancer AI Accelerator. Participants were highly engaged throughout the three meeting days, and their efforts resulted in the production of a vision and mission statement for a Cancer AI Accelerator; a draft governance structure; six focus areas or "strategic pillars"; and incentives for participation. The discussions at the meeting reflected broad support for the concept of a Cancer AI Accelerator to advance this dynamic field.

This whitepaper documents the feedback and recommendations gathered through these visioning sessions. This document forms the basis of a Cancer AI Accelerator Action Plan, under development by the trans-NCI AI Working Group[1], that will serve as a roadmap for designing and implementing this initiative. These activities are expected to include engagement with a larger community of stakeholders to define and prioritize key starting points, develop a governance structure, and launch the Accelerator with initial community-defined activities.

It should be noted that AI and its application to cancer/biomedical research is an incredibly fast-moving field. Therefore, this whitepaper is intended to represent a snapshot in time, and it is recognized that plans for the Accelerator will need to

[1] The NCI AI Working Group members are: Natalie Abrams, Oliver Bogler, Peter Choyke, Jennifer Couch, Kelly Crotty, Jonas De Almeida, Janet Eary, Emily Greenspan, Sean Hanlon, Elizabeth Hsu, Roxanne Jensen, Jerry Li, David Miller, Catherine Schweppe, Amanda Skarlupka, Umit Topaloglu, Michele Vos, Dana Wolff-Hughes, Yantian Zhang

continuously be updated and adapted based on the current and projected state of science, technology, and policy.

Of particular interest, in March 2023 more than 1,000 technology leaders and researchers wrote an open letter urging AI labs to immediately pause training of powerful large language models for at least 6 months so that a set a shared safety protocols involving enhanced accuracy, safety, interpretability, transparency, robustness, and trustworthiness can be implemented by AI developers and policymakers. This letter was followed in May by a statement released by the heads of OpenAI, Google DeepMind, and Anthropic that simply stated, “Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks, such as pandemics and nuclear war.” Although application of AI to cancer and biomedical research may be too immature to pose an existential threat, the authors of this whitepaper recognize the near-term problems including bias, privacy preservation, incorrect predictions, validation, explainability, and uncertainty that must be dealt with head-on. At a top level, it is our view that AI applied to cancer research and healthcare can be a force for good if these recognized problems are addressed and mitigated. The Accelerator is intentionally positioned to create a community framework prioritizing integrity, regulatory considerations, appropriate training, evaluation of methods, and open collaboration.

II. Overview of the Visioning Session: Purpose, Goals, and Structure

Artificial Intelligence (AI) is a term used to describe a broad category of computational and machine learning approaches that can be applied across the spectrum of cancer research and clinical practice including identifying patterns important for cancer classification, detection, diagnosis, treatment optimization, as well as mechanistic insight. For this reason, AI research is supported through a wide variety of grant and contract programs across NCI and elements of AI are embedded in numerous NCI-supported programs as the cancer research community increasingly incorporates AI methods into both discovery and translational research as well as clinical practice. While it is important to continue to develop and apply AI in collaboration with scientists and clinical specialists, there is a need and an opportunity to provide a resource for AI researchers, cancer researchers, and clinicians to share methods, resources, and approaches to understand the nuances in the development of the methods, their ethical context and the need for targeted workforce development and retention.

To enable the community of cancer-focused AI researchers and practitioners to come together, the NCI is exploring the concept of a “Cancer AI Accelerator,” a flexible, nimble, and adaptive organization to enable researchers and domain experts to develop and find appropriate methods and data, identify appropriate collaborators, work on projects of shared interest, share resources and methods, and develop and share standards as they evolve. This concept has been inspired by feedback from stakeholders working on AI in cancer research who have shared some of the challenges and needs for advancing this field. These discussions have converged on the need to foster a collective, transdisciplinary community of those developing and applying AI methods in cancer research and care.

To begin to articulate a vision for this emerging community resource, thought leaders in cancer AI research were invited to participate in a series of three working meetings to envision what could be the core structure, scope, and priorities for a Cancer AI Accelerator. Participants were engaged in a series of highly interactive activities, accomplished in working groups, toward the following outputs:

- An agreed upon Mission/Vision Statement with Guiding principles.
- A list of constituent groups to approach and some principles or guidelines for recruitment and retention of members.
- An organizational chart with draft governance structure that defines the structure of the "Cancer AI Accelerator" with defined roles, accountabilities and decision-making responsibilities- for all key stakeholders
- A strategic map with 3-6 key pillars of focus and a handful of defined measurable milestones for each pillar for the next 3-5 years.
- The beginnings of a short-term action plan for taking the Cancer AI Accelerator forward.

See Appendix A for the participant list and Appendix B for the detailed meeting agenda.

III. Vision, Mission, and Principles of the Cancer AI Accelerator

To better understand the community need and opportunity, the workshops focused on a community-defined Vision and Mission of the Cancer AI Accelerator. The participants developed a bold vision for AI in cancer research with a mission and vision for the accelerator that would enable this broader roadmap for AI in cancer and key initiatives they felt would accelerate ethical, effective use of AI in cancer research.

A. Vision

A dramatic improvement in cancer outcomes for everyone through the ethical, equitable and widespread adoption of AI.

B. Mission

To transform our understanding and treatment of cancer through the equitable development and implementation of AI by supporting the following initiatives:

People: At the heart of the mission of the Cancer AI Accelerator is a more intentional, inclusive recruitment of the next generation cancer AI workforce and the promotion of access to training and collaboration opportunities to a broader community of stakeholders. The Cancer AI Accelerator will support training to improve multidisciplinary participation and increase AI fluency.

Data: To discuss and accelerate emerging best practices for creation of AI-ready data and support the generation of more targeted data acquisition. To bring together stakeholders and foster a collaborative culture of sharing and access, of data among institutions towards a common purpose.

Models: To guide the development of reliable, equitable, and effective algorithms and minimize the risk of harm from their use or misuse. The Cancer AI Accelerator community will collaborate to develop standards and best practices for annotating, sharing and validating models.

Computation: To facilitate access to and use of efficient, sustainable, and secure computing resources that can support the rapid growth and new technologies associated with cancer AI research.

Implementation: To create guidance, tools, and support to ensure that models which are developed to advance cancer care, understanding, and innovation can be deployed across resource settings in an equitable and effective manner.

Outreach: To share materials, approaches and resources developed in the cancer AI community to broaden the impact of cancer AI, empower a broader community of stakeholders including patients and community members impacted by AI research.

C. Principles

In preparation for the Visioning meetings, NCI staff drafted a set of principles for the Cancer AI Accelerator. As the accelerator discussion progressed, these same principles arose in several discussions, building on and validating their significance:

Inclusive: Bringing varied perspectives to cancer-focused AI, developing the field through diverse contributions.

Ethical: Focused on the integrity, regulatory, and training considerations that ensure the appropriate application of AI.

Dynamic: Nimbly identify and explore evolving opportunities- It's a place where the cancer AI community identifies the most pressing areas of focus and leads the development of new initiatives to address these opportunities.

Enabling and Empowering: Provides mentorship and leadership opportunities to foster the next generation of cancer AI researchers.

Cancer-focused: Priorities are defined by driving cancer research questions.

Community-Driven: While NCI will provide the genesis, the Cancer AI Accelerator will be ongoing and community-governed.

IV. Strategic Pillars

Participants in the Visioning Sessions were engaged in distinct, small group activities that independently converged on a common set of driving considerations for the Cancer AI Accelerator referred to here as "Strategic Pillars": Data, Models, Computation, Implementation, People, and Outreach. Groups working on the Vision and Mission arrived at these pillars from "top down" conversations about the overall goals of the Accelerator. These same pillars emerged from the "bottom up" exercise of categorizing the outputs from brainstorming sessions about activities of the Accelerator. Themes that cut across all these Pillars are ethics, democratization, and identifying incentives for participation. Recommendations related to each of the pillars are summarized in the following sections, along with areas where additional input is needed.

A. Data

Data needs for advancing AI in cancer research and care received the highest volume of ideas from participants in the visioning sessions. While large data generating

activities are out of scope for the Accelerator, the participants identified collaborative opportunities to advance the availability and utility of data for training and validating AI models. Key themes are summarized below:

Improve access to and use of electronic health record (EHR) data to enable AI. Advancing the appropriate use of EHR data to train AI models was called out as an incentive for participating in the Cancer AI Accelerator. It was suggested that the Accelerator engage in community efforts to (a) break down barriers for sharing data between different health systems (federated learning and more), (b) substantially improve the accuracy of data extraction and capture in clinical settings (engage leading clinicians and medical schools), and (c) define criteria for AI-ready EHR data in terms of differing data standards, annotations, and missingness across systems and records.[2] The Accelerator should promote the availability of patient data with good annotation with respect to treatment and outcome that is adequately consented, and ensure patient protection especially for rarer cancer types.[3]

Openly share multimodal datasets suitable for development of discovery and clinical models. In several independent discussions, participants highlighted the need for publicly available, deeply annotated, multimodal datasets (with genomic, pathology, and imaging data) with associated patient outcomes and treatment response. Such datasets would be valuable for training, model testing/validation, and use in education and training. A starting point could be a community-prepared pilot version of an AI-ready multimodal dataset. This could lead to a standardized approach to data harmonization and quality assessment and enable more efficient and goal-specific model development, testing, and robust deployment.

Create a system for more universal federated learning of data/algorithms. The promise of federated learning to address the need for large, diverse, privacy-protected datasets was called out in several discussions. Specific suggestions included creation of a working group to assess considerations for applying federated learning. The working group might also identify mechanisms for federated access to clinical datasets and ultimately negotiate with health care systems to permit federated learning on clinical datasets and promote the implementation of the appropriate application of these approaches.

[2] EHR considerations are not just about getting larger and potentially more accurate data. It is also detailing with different data structures, and lots of missingness in a dataset essentially set up for administrative billing.

[3] An additional barrier is that institutions increasingly see their data as IP.

Streamline access to clinical trial data. Participants recommended that the Accelerator spearhead efforts to facilitate access to data from NCI Cooperative Groups such as NCI Navigate, CTEP, NCTN. It was also recommended that the Accelerator connect with pharma leadership to promote access to their clinical trial datasets for AI training/validation. They also suggested engaging and incentivizing industry partners in precision medicine (e.g., Flatiron, Tempus) to contribute data from failed trials. Data aggregation companies may also provide helpful input on best practices for data standardization and curation.

B. Models

Recommendations to advance the quality and translation of AI models was a critical topic in the visioning meetings. Participants were encouraged to identify activities and opportunities that would broadly advance the field rather than focus on development of any specific models. Key recommendations are summarized below.

Develop community standards for reporting on and managing AI models. Standards are necessary to promote appropriate model sharing and reuse. One specific approach called out was the development of model cards for standard model representation, including inputs and outputs; model purpose; assumptions; data source; and validation metrics. Sharing both single data modality models (e.g., imaging, genomics) and multi-modal models was encouraged. It was noted that sharing of large language models (LLMs) trained using EHR data is especially tricky given the risk of re-identification of PHI-containing datasets and resulting security barriers.

Develop an index/repository of existing research and clinical algorithms and software. Also related to model sharing, was the recommendation to create a platform to share existing research algorithms and clinical algorithms. The availability of a community-developed approach for national/global interoperability and findability across algorithms and software was seen as a key incentive for participation in the Accelerator. The analogy was drawn to NLM's PubMed for journal articles. This repository could be connected to HPC/cluster compute resources and have a mechanism for outcome reporting (performance evaluation), model comparison, validation, and generalizability assessment. The availability of a national pool of algorithms and software would help democratize access to these resources and mitigate the risk of "the rich getting richer" in cancer AI research.

Understand and lower barriers to the clinical translation of cancer AI. As stated by one participant, "*There is a big gap and effort for taking an algorithm from development to clinical implementation. It would be great if the Accelerator could help with this."* Participants suggested that the Accelerator develop pathways to translation of robust,

highly predictive, validated models to the clinic through activities such as matchmaking with clinical care teams and clinical trials. There is also the need to identify what performance measures are clinically important and long term management of these models in a clinical setting.

Foster development of a standard framework and platform for development, validation, consumption, and implementation of AI algorithms. Participants highlighted the need to support the full life cycle of an AI algorithm, which includes post market surveillance and societal implications of AI algorithms. This could include defining a framework for maturing algorithms from definition of a problem through to production and dissemination. They recommended creation of a federated/collaborative evaluation and testing infrastructure to compare local models vs. global models vs. federated learning.

Interpretable and explainable AI models. Participants emphasized the need to invest in model interpretability and explainability. Understanding the reasoning behind predictions made by a model vs. "black box" models that have no explanation of why an AI arrived at a specific decision is crucial in building trust and comprehension of AI-based approaches. Explainability helps characterize model accuracy, fairness, transparency, and outcomes in AI-powered decision making.

Foundation models. Foundation models are very large neural networks that are trained on huge amounts of data. Cancer AI researchers can apply pre-trained Foundation models to a wide variety of downstream tasks, rather than building individual AI models. This results in greater model reuse and sustainability and lower cost and time investment.

C. Computation

A common theme throughout the Visioning meetings was the need to broaden and democratize access to the computing resources necessary to train robust AI models. As one participant noted, "*Large language models reach 500 billion parameters, running them requires non-trivial compute resources for inference*." It was noted that academic labs are at a severe disadvantage compared to industry. A single lab does not have the resources to compete. A long-term vision shared by one participant is a sustainable, secure, and energy-efficient infrastructure to handle increasing biomedical data and AI computation. Key recommendations are summarized below:

Create a comprehensive catalog of cancer/biomedical advanced computing/HPC capabilities and how they are used and accessed. Participants noted that adoption of cloud computing does not always meet the technical or financial needs of academic

cancer AI researchers as new technologies and methods, such as Foundation models and field programmable gate arrays (FPGAs), are developed at a rapid pace. Gaining a comprehensive understanding of both cloud and local HPC capabilities available to cancer AI researchers, along with specific use cases would be a valuable resource.

Provide nation-wide coordination of computing management, operation, and infrastructure support. Participants recommended sharing computing resources with NSF, DoE or creating an NCI network of computing resources following NSF models. A related suggestion was to create a program to encourage biomedical, HPC, and CS researchers to make coordinated requests for high end and leadership scale compute resources.

Develop and deliver training for HPC / Cloud for researchers at all levels. One specific suggestion was to create a best practice guide for using cloud vs local compute resources. Additionally, workshops on topics such as when to use GPU vs CPU were also recommended.

D. Implementation

Recommendations for the Implementation Strategic Pillar have interdependencies with the Data, Models, and Compute recommendations. Ethics, democratization of resources, and identifying incentives are also important themes inherent to implementation. Key recommendations are summarized below.

Create standardized templates for commonly used processes. Participants indicated that investigators commonly struggle with important legal and administrative processes related to data use and data sharing. There is an opportunity for the Accelerator community to create agreed-upon, standardized, templates for commonly used documents such as data usage agreements (DUAs), IRB submissions, Statistical Analysis Plans (SAPs) and IP agreements that can be adopted and shared broadly, especially as it relates to de-identified data and data re-use. A similar idea relates to creating templates or trusted sources for technological needs, such as de-identification, quality assessment, data assimilation across different modalities, orchestration of docker containers, model portability, and uncertainty quantification. It might also be advantageous to have the Accelerator support “core labs” to standardize and perform common technological/engineering tasks.

Clinical implementation of AI. Challenges related to the clinical implementation of AI models were a significant topic of discussion across the Visioning meetings. Participants pointed out that it is very difficult to get funding support for clinical deployment/implementation of AI since it is not hypothesis driven research. They

recommended tailored funding opportunities to study the process of implementation, validation, deployment, and how humans interact with these systems. It would also be advantageous to develop a separate community white paper to articulate the need for AI software clinical validation. More focus within cooperative groups to implement AI-driven prospective clinical trials was also suggested, as well as opportunities to develop secondary endpoints based on AI-driven analyses. AI secondary endpoints. As a general theme, it was also suggested to leverage existing efforts whenever possible, for instance, by collaborating with existing companies offering “AI Orchestration” services to facilitate clinical implementation and testing.

Training opportunities for AI implementation: Near-term opportunities to advance topics related to AI implementation include organizing workshops or educational sessions on topics such as translating research code to production ready code, quality management systems and code documentation.

E. People

For the “People” Strategic Pillar, participants identified the following overall goals: 1) Develop a community of practitioners; 2) Promote training awareness; and 3) Support a sustainable workforce. Integral to all aspects of this Strategic Pillar is facilitating diversity, equity, and inclusion in the cancer AI workforce. Key recommendations are summarized below.

Support for junior investigators and training. Participants noted the importance of mechanisms by which junior investigators can access clinicians and technical researchers with important unsolved problems and seek guidance/mentorship to guide career activities/goals. Training may take multiple forms, including hands-on workshops across different disciplines on various scientific, computational and organizational (e.g., how to develop productive collaborations) topics. There is also a need for training grants and early career investigator awards focused on cancer AI research. Emphasis should be placed on involving investigators from diverse demographic and educational backgrounds and providing support for investigators from underrepresented institutions. The challenge of salary differences between academia and industry was also noted.

Develop a system for expert “office hours”. There is a desire among the community to develop a structured knowledge-sharing program where individuals can seek guidance on common issues and roadblocks to research progress. Related to this is a desire to support peer-peer matchmaking that would allow investigators at all career stages to receive feedback on their research ideas and identify collaborators across different disciplines.

Support clinical collaboration. The Accelerator should engage clinicians to advance clinical implementation of AI methods. This could include involving clinicians in research design and intentionally developing AI methodologies and validation with an end goal of clinical application/clinical trials. Participants indicated a need for alignment of clinical and research incentives for long-term sustainability of method development. It will also be important to emphasize training MDs interested in AI research and/or applications of AI/ML to their clinical practice. A specific idea is to focus the training on certificate courses that would confer "AI Literacy" to critically appreciate and understand vendor offerings.

F. Outreach

To identify partnering needs and opportunities, participants were asked to make a list of organizations, associations, companies, and programs whose work aligned with the mission of the AI Accelerator. In this exercise, participants identified dozens of such entities. The complete list of groups is provided in Appendix C. Priorities identified by the group included:

- **Federal agencies**: Including agencies with related initiatives such as NLM and NSF
- **Integrated medical networks**: Including the VA and Kaiser, with access to large datasets and parallel ongoing efforts in AI
- **Cancer registries**: Such as SEER, with its mass aggregation of patient data
- **Professional societies**: Including ASCO and AHS, with partnering opportunities that may include framing policy for AI in cancer
- **Large data generators**: Including Roche/Flatiron for potential access to clinical trial datasets and for framing clinical questions around unmet needs for AI (e.g., companion diagnostics)

There is also broad interest in partnering with venture capital (VC) firms to accelerate cancer AI technologies towards commercialization. The Accelerator could support a community-developed framework for constructive outreach and engagement among industry, VCs, and NCI with academic leaders to prioritize the key questions, opportunities and needs where AI would result in the greatest impact in the short and long term. Potentially, the Accelerator could partner with VCs to screen ideas that could then be accelerated commercially.

Constructive engagement with the media about AI applications in cancer will be important to avoid overhyping new developments, setting appropriate expectations about technologies, using popular science language in engagement, and working to educate the media about AI/ML in cancer. Members of the Accelerator could both engage in this outreach and support media training for investigators.

Some participants urged consideration of the roles of cancer patients, patient advocates, caregivers/care partners, and the public in the Cancer AI Accelerator program.

V. Motivating and Sustaining the Accelerator

At the end of the AI Accelerator Visioning meetings, participants were asked to consider community motivations and incentives for participating in a Cancer AI Accelerator. For this exercise, two subgroups were asked to brainstorm the "top 10" incentives for Senior Scientists and two other subgroups were asked to do the same for Junior Scientists. The ideas from these exercises have been synthesized and summarized below:

1. **Improved dataset creation, access, and opportunities for use**
    a. Creation of a publicly available, deeply clinically annotated, multimodal dataset with associated patient outcome, treatment response. Provide tiered access to data with respect to privacy as well as appropriate tools and data structures.
    b. Promoting public access to clinic trial datasets from oncology cooperative groups and pharma companies.
    c. Organize challenges involving the use of well-annotated datasets. Involve industry partners to facilitate integration as the final deliverable.
    d. Generating a standardized approach to data harmonization and quality assessment to enable more efficient and goal-specific model development, testing and robust deployment.
    e. Make available useable real-world clinical records for AI learning and prediction by engaging in community efforts to (a) break down barriers for sharing data between different health systems[4] (federated AI learning and more) and (b) substantially improve the accuracy of data extraction and capture in clinical settings (engage leading clinicians and medical schools)

2. **Sharing and documentation of AI/ML/Statistical models**
    a. Community consensus in model sharing. This includes models ranging from single data modality models (e.g. imaging, genomics) to multi-modal models as well as large language models (LLMs) built from EMR data

[4] Example: Make a HIPPA compliant, multi-institution HL7 feed from a common EHR that could be used as a training set for ML and training

(taking into account the consideration of potential memorization by the model of PHI and other legal barriers).
   b. Community-developed approach for national/global interoperability and findability of algorithms, datasets, and tools.

3. **Facilitation of collaboration and addressing important problems**
   a. Articulate a set of well-formulated questions that motivate the advancement of the technology to result in clinical impact and encourage interdisciplinary groups to form and work on these questions.
   b. Match-finding mechanism for collaborators with different expertise. For example, access to clinicians and researchers with important and unsolved problems and/or access to industry partners..
   c. Nationally recognized cooperative group providing potential for collaborations, access to clinical experts, reliable access to accumulating data, and credible governance to recognize efforts and academic products/models/scholarly work. A place to be!
   d. The excitement of solving a technical problem that could improve hundreds or thousands of lives.

4. **Documented and shareable best practices that evolve with time**
   a. Creation of community-established standards for model evaluation, effectiveness, robustness, deployment, and outcome reporting to appropriately address regulatory requirements.

5. **Develop strategy to substantially lower compute cost**
   a. Partner with cloud providers
   b. Best practice guide for using cloud vs on prem compute – Potential compute approaches, e.g. sharing computing resources with NSF, DoE or creating NCI network of computing resources following NSF models.
   c. Access to affordable high performance computing resources that are paired with medical data access.

6. **Education and training that promote cross expertise and varied career goals**
   a. Novel training programs: Develop new graduate training programs for PhD, MD/PhD, and postdocs to encourage new people to enter the field and develop expertise. Increase focus on deep cross-training.
   b. Novel educational opportunities: For example, provide “Training up” workshops for PhD in innovative/emerging methods to teach the wider community. Workshops to train early stage investigators and foster ideas /

inter-institutional collaborations. Education for MD/clinical in the opportunities/techniques of AI etc. MDs who want to take advantage of AI; Technical people who want to understand the clinic.

c. Guidance, mentorship, and best practices to guide career activities/goal. This could include VC-like pitch competition paired with availability to entrepreneurship mentoring.

7. **Easing logistical constraints and obstacles for early career scientists**
    a. Funding opportunities, commercial and federal; early career funding
    b. Enabling junior investigators to take risks with protection from negative impact from failure/ lack of generality / failure to translate, etc.
    c. Wiki/blog-like guidance for tasks/obstacles to building/deploying, processes, etc. Standards for model deployment framework for maturing models.
    d. Public guidance to standardize credit for multi-author AI efforts, including recognition for peer-reviewed proceedings. Dedicate funding for “high quality data generators.”
    e. Competitive salaries (especially for postdoctoral fellows, given increasing opportunities in industry)

8. **Community activities to address challenges related to clinical implementation**
    a. Integrating uncertainties in model development and output representation to refine human/AI interaction (e.g.: interpretability, trust, credibility)
    b. Ensuring safeguards (to increase patient safety and accurate outcomes).
    c. Pathways to translation of robust, successful, validated models to the clinic including matchmaking with clinical teams[5]

9. **Aligning of clinical and research incentives**
    a. Healthcare systems are increasingly interested in operationalizing AI and rely on senior faculty heavily in these activities. This may involve use of NCI-funded research tools and present opportunities for research on the real-world value of AI. Funding opportunities or support for these activities or realigning the scope of FOAs could be an incentive for senior faculty.

[5] It was noted that much of this will be opportunistic (does the data already exist) so perhaps there is some way to 'train' people to match existing data with important questions.

VI. Governance of the Cancer AI Accelerator

A. Governance Structure

Through a series of facilitated exercises, participants were asked to consider the organizational elements and governance structure for the Cancer AI Accelerator. These discussions resulted in organizational structures focused around leadership, advisory/oversight groups, and working groups.

Leadership: Both groups recommended a single scientific lead for the Accelerator with term limits. One group additionally recommended an Operational lead that would provide continuity.

Advisory/oversight groups: There was consensus on having additional advisory/oversight groups. These groups could vary along several axes, including composition (academic, NCI, industry), reporting structure (e.g., reporting to NCI or another group in Accelerator), and role (e.g., big picture directions vs. technical, scientific aspects). Potential structures/terminology proposed include:

- Steering Committee/Executive Committee: Eight to ten rotating members of the community that appoints/selects working groups.

- Advisory Council/Board of Directors: Stakeholders represented at a high level that can position the accelerator nationally/internationally. Also seen as advising the Steering Committee.

- Working groups**:** Beyond leadership and advisory groups, there was a consensus on the need for working groups that would focus on executing specific initiatives with well-defined endpoints.

B. Overarching Principles of the Cancer AI Accelerator

Overarching principles frequently brought up in the context of structure and governance included **diversity, flexibility**, **efficiency**, **community-led**, and leveraging **pre-existing knowledge and established organizations**. There was broad consensus on avoiding a strongly top-down structure and there was an emphasis on avoiding redundancy with existing initiatives and instead leveraging these efforts to achieve new goals.

C. Post-meeting reflections and gaps

A question posed astutely during the visioning session: “Where do we begin?” The process and timeline for establishing essential governance components has yet to be determined and is a critical first step. Which pieces of the organizational structure are needed first? Would it be possible to build the structure out over time, or would the Accelerator benefit from a robust organizational structure at initiation? The identification of the scientific lead and/or and oversight board might benefit from input from the broader Cancer AI research community.

VII. Membership in the Cancer AI Accelerator

A. Summary of recommendations

The Cancer AI Accelerator is envisioned to be an inclusive community of researchers committed to advancing cancer research and improving cancer outcomes through the application of advanced AI methods. A fundamental consideration is defining who/what is meant by “community”? Who are the members of this community? How can we broaden participation? Meeting participants were asked to suggest membership requirements, processes to become a member, and member responsibilities.

Membership requirements: Participants recommended an open, inclusive approach to membership. An interested researcher would sign up to create a simple profile indicating interests that are searchable in a directory of members and would be “admitted” after signing a Code of Conduct. The group acknowledged that specific initiatives within the Accelerator may have additional requirements or restrictions for participation. A Stack Overflow-type of online community was suggested to build a social network in which to ask and answer questions among members of the community.

Membership categories: The participants proposed a distinction between “Active Activities” and “Observer Activities”. “Active Activities” would correspond to specific initiatives such as standard setting, supplying data, assembling teams to address wish list items, organizing platforms for communication, etc. “Observer Activities” are for members not currently involved in “Active Activities,” including observing/evaluating activities in the communication platform, disseminating information to others, and submitting a monthly wish list (e.g., membership or other specific needs). Members

would not be removed for inactivity; however, activity would increase a member's role in the community.

Member activities: Governance would provide a structure with milestones and tasks to rally members to get involved. With this structure, there would be organic formation of teams in which each member would play a role in creating tangible outcomes in a short time period (e.g., 3 months) within a larger overarching plan. Tasks would be carried out by teams of 4-5 people that should be equitable in terms of seniority and experience but without specific stipulations or requirements. It was suggested that regular turnover be promoted but not required on these teams, in an environment that facilitates fresh thoughts, collaboration, and new ideas.

B. Post-meeting considerations

Outstanding membership questions and suggestions that would benefit from further discussion with the broader community include:

1. *Should we refer to "opportunities for participation" rather than having the concept of "membership"?*

2. *Should there be an application process for membership to the Accelerator?* The rationale for an application process is that raising the bar, even slightly, for participation in the Accelerator would help ensure that people who join the Accelerator feel some level of commitment to the program. For example, people could be asked to provide a short paragraph describing what they hope to get out of participating in the Accelerator. Most applications could be approved, with the idea that even applying is a sufficient "bar".

3. *How do we address diversity, inclusivity, and democratization in the Cancer AI Accelerator*? An important, recurring theme in the Visioning discussions was the need for the Accelerator to provide an inclusive environment that fosters participation of a diverse community of researchers. To truly accomplish this goal will require a very purposeful approach. Engaging an expert in building inclusive environments should be considered.

VIII. Next Steps

The next steps to advancing the Cancer AI Accelerator are to develop an Accelerator Action Plan; engage with a larger community of stakeholders for input and support,

identify a Community Manager, and launch the Accelerator with initial community-defined activities.

Develop an AI Accelerator Action Plan: NCI staff will draft an AI Accelerator Action Plan that will include initial recommendations for the community structure and governance that is consistent with the outputs from the visioning sessions as well as plans for monitoring progress and evaluating impact.

Convene the broader cancer AI community to gain additional input to develop the Accelerator and to identify the early activities. Host a public event to engage a much broader audience in developing and launching the Accelerator. This community meeting would be preceded by a series of “microlabs” – small, virtual events to share the vision and draft Action Plan for the AI Accelerator and build interest in the concept. The goals for this event are to attract participants to the Accelerator, identify the inaugural working groups, and nominate initial projects and initiatives.

Identify a Community Manager for the Cancer AI Accelerator. Input from the Visioning meetings validated the need for an operational lead for the Accelerator, to provide consistent management and coordination of this community.

Launch the Cancer AI Accelerator. For the Cancer AI Accelerator to gain momentum, it will be important to kick off one or more of the activities identified in the community meeting as quickly as possible. Based on some of the ideas from the Visioning meetings, compelling activities might include:

- An Innovation Lab framed around a community-defined topic, where participants would creatively explore new ideas, propose pilot projects, and compete for prize funding to support the project.
- Prototype a mentorship/office hours activity to connect senior researchers with junior researchers in cancer AI.
- Collect a list of existing training materials and programs like summer student programs to share with the community.

IX. Appendix

A. Participant List

Last Name	First Name	Institution
Basu	Amrita	UCSF
Biven	Laura	NIH

Brown	Sherry-Ann	Medical College of Wisconsin
Cooper	Lee	Northwestern
Davidson	Natalie	CU Anschutz
Deasy	Joseph	MSKCC
Deng	Jun	Yale
Elemento	Olivier	Cornell
Fraenkel	Ernest	MIT
Gevaert	Olivier	Stanford
Gibbons	Chris	MD Anderson Cancer Center
Greene	Casey	U Colorado
Hernandez-Boussard	Tina	Stanford
Ideker	Trey	UCSD
Jaffray	David	MD Anderson
Janowczyk	Andrew	Case Western
Kalpathy-Cramer	Jayashree	MGH
Kibbe	Warren	Duke
Kozloski	James	IBM Research
Lotter	Bill	Harvard
Macklin	Paul	Indiana University
Madabhushi	Anant	Emory
Mesirov	Jill	UCSD
Parikh	Ravi	University of Pennsylvania
Peng	Grace	NIH
Pomann	Gina-Maria	Duke
Quackenbush	John	HSPH
Raphael	Ben	Princeton
Saltz	Joel	Stony Brook
Sanders	Chris	Dana Farber
Savova	Guergana	Harvard
Siegel	Eliott	University of Maryland
Simpson	Amber	Queens University
Tourassi	Gina	ORNL
Yetisgen	Meliha	University of Washington

NCI Organizers
Jennifer Couch, DCB
Emily Greenspan, CBIIT
Sean Hanlon, CSSI
Juli Klemm, CSSI
Roxanne Jensen, DCCPS
Catherine Schweppe, NCI Presidential Management Fellow

KnowInnovation Facilitators

Tim Dunne
Zach Stapleton Jones

B. Detailed Agenda of the Visioning Sessions

Day 1 – Tuesday, January 10, 2023

- Welcome from NCI
- "What Could Go Wrong": Small group activity – What are all the things we want to make sure don't happen as we build the Cancer AI Accelerator?
- Report Backs
- Mission and Vision Feedback: Small group activity – Feedback and suggested additions, deletions, and edits to the draft Mission Statement
- "What's Lacking?": Small group activity – Get a list of all the challenges, problems, things that are lacking that a Cancer AI Accelerator might address using a Mural board
- "Moving Forward": Three simultaneous small group activities
    - Refining the Mission Statement
    - Parsing the "What Could Go Wrong" activity
    - Making sense of the "What's Lacking" activity

Day 2 – Tuesday, January 17, 2023

- Mission and Vision Review: Full group discussion about the Mission and Vision after hearing from Chris Gibbons
- Strategy: Small group activities – Choose between
    - Building on the Mission and Vision
    - Map the opportunities from the Mural board activity into a roadmap for each Strategic Pillar: People, Data, Compute, Models, Implementation, Outreach
- Governance: Small group activity – Create governance and organizational structure using a metaphor – House, Cell, Computer, Recipe, Town
- Report backs

Day 3 – Thursday, January 26, 2023

- Primer activity: Full group activity Look for milestones, resources, outputs, or services that would be useful to you, are exciting, or would make you want to engage with the AI Accelerator.
- Sustaining a Community: Small group activity exploring incentives for participating in the AI Accelerator for Senior Scientists and Junior Scientists. Each group was asked to create a Top 10 list and consider:
    - What might be all the ways to incent and sustain participation in the AI Accelerator - from the perspective of the work the Accelerator is doing?

        - What would make you want to participate?
        - What will make this 'place' useful to you (think resources)?
        - What is the work that the community could do?
        - If we have the community engaged, and pilot level funding how could you advance the field in ways that you couldn't otherwise?
        - Which of the Mural milestones are exciting to you, act as a pull? Think about what can this community do better than other vehicles.
- Report backs
- Structure: Small group activities
    - Two groups built out an Organizational and Governance Structure and roles and responsibilities
    - Two groups thought about the structure of the membership, with the following prompts
        - What are the membership qualifications or requirements? (Is membership even the right word/concept?
        - What is the process for how to become a 'member', including any formal agreements?
        - What are the benefits of membership?
        - What are the major responsibilities of a member?
        - What are the stipulations of membership (e.g. tiers, term limits, etc.)?
        - Could there be different types of memberships for different groups?
        - Is there a cap or threshold on the number of members?
        - Is it possible to lose membership, and if so what are the metrics and process?
    - One group considered the landscape (or ecosystem) of organizations and contributors outside of, but impacting, the AI Accelerator network
- Open Discussion: Full group time to reflect on our progress and look to the future

C. Organizations Engaged in Activities Related to the Mission of the Cancer AI Accelerator

Societies

Digital Pathology Association (DPA) - Professional Organization focused on digital pathology, mostly vendor based, big focus on AI in pathology

Association for Pathology Informatics (API) - similar to DPA, digital pathology focus

Radiological Society of North America (RSNA)

ABAIM: American Board of AI in Medicine (https://abaim.org/leadership) – Mission: Formed by a team of leading clinicians and data scientists, The American Board of

Artificial Intelligence in Medicine (ABAIM) has emerged as the authoritative body for medical AI education and certification.
MICCAI - Medical Image Computing and Computer Assisted Intervention
American Medical Informatics Association (AMIA) – the premier organization of medical informaticians
ASCO - American Society of Clinical Oncology
SITC - Society for Immunotherapy in Cancer
AACR - American Association for Cancer Research
ESMO - European Society for Medical Oncology
Human Immunome Project (HIP)

Federal
National AI Initiative
VA
National Library of Medicine – very similar initiatives, funding lots of AI for biomedicine in general.
Most academic institutions have AI or AI-adjacent programs, they contribute to different facets of AI (from theory to applied)
National Institute of Biomedical Imaging and Bioengineering - funds a lot of AI research mostly applied to imaging
NHGRI
Veterans Affairs Medical Center (e.g. LPOP - Lung Precision Oncology Program)
National Science Foundation (NSF)
DARPA/CDMRP
Frederick National Lab
Cancer registries, e.g. SEER
CDC and the CDC cancer registries
PCORI - Patient Centered Outcomes Research Institute
Oncology Cooperative Groups - SWOG, ECOG-ACRIN, NRG
Cancer Imaging Archive (TCIA)
Imaging Data Commons (IDC)
MVP - Million Veteran Program (VA),All
All of Us
EU **Cancer** Diagnostic and Treatment for All

Academic Centers
Eric and Wendy Schmidt Center at Broad Institute (250M endowed center focused on the intersection of AI and the biomedical sciences)
Gladstone Institutes: Biomedical research institute with one of the pillars being on AI
CI4CC

NCI Designated Comprehensive Cancer Centers

For Profit/Commercial
Sequencing providers, e.g. Tempus - an AI focused data-aggregator company that provides sequencing services for oncology; Foundation Medicine, Cares
Pharma (past successful/productive collaborations) - BMS, ImmunAI, Astrazeneca, Eli-Lilly, Janssen/J&J, Sanofi, Roche/Flatiron/GenenTech