The Stakeholder Playbook for Explaining AI Systems

Robert R. Hoffman
Institute for Human and Machine Cognition

Gary Klein
Macrocognition, LLC

Shane T. Mueller
Michigan Technological University

Mohammadreza Jalaeian
Macrocognition, LLC

Connor Tate
Institute for Human and Machine Cognition

Keywords: explainable AI, stakeholders, requirements

This material is approved for public release. Distribution is unlimited. This material is based on research sponsored by the Air Force Research Lab (AFRL) under agreement number FA8650-17-2-7711. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL or the U.S. Government.

Cite as:
Hoffman, R.R., Klein, G., Mueller, S.T., Jalaeian, M., and Tate, C. (2021). " The Stakeholder Playbook for Explaining AI Systems." Technical Report, DARPA Explainable AI Program.

Abstract

The purpose of the Stakeholder Playbook is to enable system developers to take into account the different ways in which stakeholders need to "look inside" of the AI/XAI systems. Recent work on Explainable AI has mapped stakeholder categories onto explanation requirements. While most of these mappings seem reasonable, they have been largely speculative. We investigated these matters empirically. We conducted interviews with senior and mid-career professionals possessing post-graduate degrees who had experience with AI and/ or autonomous systems, and who had served in a number of roles including former military, civilian scientists working for the government, scientists working in the private sector, and scientists working as independent consultants. The results show that stakeholders need access to others (e.g., trusted engineers, trusted vendors) to develop satisfying mental models of AI systems. and they need to know "how it fails" and "how it misleads" and not just "how it works." In addition, explanations need to support end-users in performing troubleshooting and maintenance activities, especially as operational situations and input data change. End-users need to be able to anticipate when the AI is approaching an edge case. Stakeholders often need to develop an understanding that enables them to explain the AI to someone else and not just satisfy their own sensemaking. We were surprised that only about half of our Interviewees said they always needed better explanations. This and other findings that are apparently paradoxical can be resolved by acknowledging that different stakeholders have different capabilities, different sensemaking requirements, and different immediate goals. In fact, the concept of "stakeholder" is misleading because the people we interviewed served in a variety of roles simultaneously — we recommend referring to these roles rather than trying to pigeonhole people into unitary categories. Different cognitive styles re another formative factor, as suggested by participant comments to the effect that they preferred to dive in and play with the system rather than being spoon-fed an explanation of how it works. These factors combine to determine what, for each given end-user, constitutes satisfactory and actionable understanding.

Outline

1. Introduction

The initial focus of much of the activity under the rubric of Explainable AI was on developing AI systems that could explain themselves to End-users. The Stakeholder Playbook was created in recognition of the possibility that stakeholders also need explanations, but also that different stakeholders would need different kinds of explanations (forms and contents) depending on their roles and responsibilities. The purpose of the Stakeholder Playbook is to enable system developers to appreciate the different ways in which stakeholders need to "look inside" the AI/XAI system. For example, some stakeholders might need to understand the boundary conditions of the system (its strengths and limitations). Program Managers might need to understand an AI/XI system in a way that enables them to succinctly explain the system to other people. Leaders of system development teams need to be able to develop appropriate optimism, informed by appropriate skepticism. By hearing first-hand from the different stakeholders about what they need in terms of explanations, developers will be better able to help stakeholders develop good mental models of a system. The Stakeholder Playbook is intended as guidance for all of the responsible stakeholders, but especially guidance to developers who are creating the explanation capabilities of their AI or XAI systems.

This article begins by encapsulating the pertinent literature, and the important considerations with regard to how to taxonomize stakeholder groups. We then present the method and results of an empirical investigation that led to the Playbook. A number of responsible individuals were engaged in cognitive interviews concerning their roles and responsibilities. Data collected through the interviews yielded a number of surprises.

2. Background

Stimulated largely by the Report of the European Union on the notion of a "right to an explanation" (European Union, 2016; see also Goodman and Flaxman, 2017; Wachter, Mittelstadt, and Floridi, 2016), a number of U.S. policy making organizations have issued policy requirements. These are listed in Table 1.

Table 1. Some U.S. policy making organizations that have put forth legal explanation requirements.

ORGANIZATION	POLICY	REQUIREMENT
US Federal Reserve	SR 11-7	Fairness, Privacy, Transparency, Explainability Model validation and Verification
European Commission	General Data Protection Regulation Article 22	Individuals have a right to demand an explanation of how an AI system made a decision that affects them
US Congress	Algorithmic Accountability Act 2019	Companies are required to provide an assessment or risks related to accuracy and fairness.
US Executive Branch	Executive Order No. 13,859, 84 Fed. Reg. 3967	Foster public trust and confidence in AI technologies; establish guidelines and standards to enable the regulation of AI technologies, with the aim of enabling innovation while protecting privacy and national security interests; reduce

		barriers to the safe testing and deployment of AI technologies
Washington State Legislature	HB 1655	Transparency to protect consumers
State of California	Consumer Privacy Act	Consumer protection
State of Illinois	HB 3415	Consumer protection

interest in XAI has expanded, so too has interest in the notion that different stakeholders will need different kinds of explanations. Awareness of the importance of this topic has led many researchers in computer science to propose mappings of stakeholder groups onto explanation requirements. We now briefly review this literature.

Recent Pertinent Literature

In this recent literature on stakeholder dependence, the analysis of explanation requirements is speculative and sometimes superficial, for example asserting that developers need "scientific" explanations whereas End-users or lay persons need "everyday" explanations. Papers in this burgeoning literature include:

- Ones that emphasize the importance of stakeholder dependence and provide a roster of different types of explanations (e.g., Dahan, 2020; Hind, et al., 2019; Floridi, et al., 2018; Preece, et al., 2018),
- Ones that emphasize the importance of explanation for the ethical implications of AI (Floridi, et al., 2018; Goodman and Flaxman, 2017),
- Ones that refer to stakeholders as a concept and emphasize the importance of transparency, but do not then consider explanations or explanation effectiveness (e.g., Eiband, et al., 2018),
- Ones that ostensibly refer to explanation and explainable AI but focus on formal interpretability or transparency rather than pursuing a notion of explanatory value (e.g., Felzmann, et al., 2019; Kaur, et al., 2020; Tjoa and Guan, 2020; Tomsett, et al., 2018), and
- Ones that discuss the matter by referencing other papers that offer speculations (e.g., Langer, et al., 2021; Mittelstadt, Russell and Wachter, 2019).

Most of the discussions of the stakeholder relativity refer to a small set of stakeholder types, often only to a distinction between end users, system developers and "others." Naiseh, Jiang, Ma, and Ali (2020) offer a broader palette of stakeholder types, integrating lists from Ribera and Lapedriza (2019), Tomsett, et al. (2018); Hind, et al., (2019): creators/developers, AI researchers, lay users, operators, domain experts, decision makers, affected parties, ethicists, theorists, external regulatory entities or oversight organizations.

Table 2 is a version of the "Users Chart" developed by DARPA (2020), which presented assertions concerning the explanation requirements of different stakeholder groups.

Table 2. A version of the "Users Chart" developed by DARPA.

Stakeholder Group	Sensemaking Needs	Explanation Requirements
Developers: AI Experts Test Operators	Does the system work well? Why do errors occur?	Explanations must be useful in refining the AI. Explanations must expose fine details.
Polity Makers Regulators	Is it fair? Does is reflect existing law?	Explanations must be defensible. Explanations must provide clear rationale.
Operations: Military, Legal, Transportation, Security, Finance, Medical	Does the AI aid decision Making?	Explanations must justify decisions and actions.

Hoffman, Klein, and Mueller (2020) added more stakeholder groups to the DARPA list (Human Factors/Cognitive Systems Engineers, Development Team Leaders, Program Managers, Procurement Officers, Systems Integrators, Vendor Managers, Trainers, Policy Makers, Regulators, Legal Practitioners, Rights Advocates, and Guardians). They also speculated about the explanation requirements associated with certain roles that had been left out of other discussions. For example, Human Factors/Cognitive Systems Engineers need to know why errors occur. They also may need explanations that expose fine details, including the rules that the AI follows.

Clearly, there is a need to look beyond the classification of "responsible individuals who need explanations" into just a few stakeholder categories, to a consideration of a significant number of roles and their combinations. For example, the list can be expanded to consider Trainers. They need to understand an AI system well enough to be able to explain it to others (trainees). On the development side, the analysis should consider Procurement Officers, Vendor Managers, and Program Managers. On the applications side, the analysis should consider Lobbyists, Legal Rights Advocates, Ethicists, Politicians, and other roles as well.

Many recent papers have offered speculations about the explanation requirements of various stakeholder groups (Floridi, et al., 2018; Goodman and Flaxman, 2017; Hind, et al., 2019; Kaur, et al., 2020; Tjoa and Guan, 2020). Arya, et al. (2019) presented a taxonomy of types of explanation types and explanation methods, and speculated on how the different types would be more helpful to loan officers versus loan applicants versus bank executives. Langer, et al. (2021) listed 28 explanation requirements such as acceptance, confidence, fairness, and performance. These are mapped onto four stakeholder types: users, developers, regulators, and affected parties. Attesting to the burgeoning interest in this topic, Langer et al. cited over 100 papers that "claim, propose, or show that XAI-related research (e.g., on explainability approaches) and its findings and outputs are central when it comes to satisfying [the explanation requirements]" (p. 7).

Hind, et al. (2019, p. 124) offered a speculative mapping of stakeholder types onto different explanation requirements: End-users (decision-makers) need explanations that let them build trust, and "possibly provide them with additional insight to improve their future decisions and

understanding of the phenomenon." Affected Parties need explanations that enable them to determine whether they have been treated fairly. Regulatory bodies need to know that decisions are fair and safe. System builders need to know if the AI is working as expected, how to diagnose and improve the AI, and possibly gain insight from its decisions. Hind, et al. asserted that all explanations of AI systems must present a justification for the AI's decisions, must contribute to user trust, must include information that the user can verify, and must rely on the domain concepts and terminology. They also asserted that the complexity of an explanation must match the "complexity capability of the user," but this interesting speculation is not analyzed.

Weller (2019, pp. 2-3) presented a mapping of stakeholder groups onto different "types of transparency," which might be understood as types of explanations. Developers need to understand how their system is working, aiming to debug or improve it, to see what is working well or badly, and get a sense for why. Users need to have a sense of what the AI is doing and why, need to feel comfortable with the AI's decision, and need to be enabled to perform some kind of action. Experts/regulators need to be able to audit a decision trail, especially when something goes wrong. Members of the general public need to feel comfortable so that they keep using the AI.

Although the majority of work in this area has centered on the taxonomics we have just reviewed, these notions have begun to make their way into products. For example, IBM's AI Explainability 360 demonstration (IBM, 2020) has distinct interfaces for three different 'consumer' types (bank customer, loan officer, and data scientist), which offer increasingly complex perspectives on decision rules as well as moving from case-based examples as explanations to large-scale views of underlying data in a loan application scenario. Although the IBM demos work primarily to suggest potential interfaces and algorithms appropriate for different stakeholders, and do not prescribe or dictate the interface, they represent concrete example of how explanations might be tailored to different stakeholders and roles.

It is safe to say that all of the researchers and scholars working in this area agree with Hind, et al., (2019) that different stakeholders likely need different explanations that match with the beneficiary and are tailored to the domain. Liao, Gruen and Miller (2020) went a step further, asserting that because different users of AI need different types of explanation, users of AI in various domains should be involved in the process of developing explanations for AI decisions in their domain.

Most papers on stakeholder-dependence do not present an analysis of explanation requirements, beyond asserting requirements and their mapping onto stakeholder groups. An exception is papers that refer to explanation in the context of legal argumentation, since such papers often have jurisprudence professionals as co-authors (e.g., Al-Abdulkarim, Atkinson, and Bench-Capon, 2016; Al-Abdulkarim, et al, 2019; 2018; Felzmann, et al., 2019; Langer, et al., 2021). In that domain, the issues of how explanations are interpreted and the effectiveness of explanations is salient. Atkinson, Bench-Capon, and Bollegala, (2021) provide a thorough review of AI applications in the legal domain, including work on the automated extraction of legal reasoning from case corpi.

The Challenge of Categorization: Groups or Roles?

Most of the work on stakeholder explanation requirements assumes that different stakeholder groups will require different kinds of explanations, differing in form and content. As we will explain, our results show that this is only true to a certain extent. Individuals who would fall into the same stakeholder group or category can nonetheless have different roles and

responsibilities (which are typically in flux) and, therefore would have different sensemaking needs and explanation requirements. The distinction between a stakeholder group or category and specific roles within categories is also important because an individual might serve in more than one role. Furthermore, the roles adopted by an individual might cut across stakeholder groups. For example, an End-user might also have the skills and motivation needed to engage in some software development. A Developer might also serve as a Trainer. This type of role expansion may make it impractical to try and aim a single specific type of explanation at a specific stakeholder group.

While an individual serving in a role might have similar explanation requirements to an individual in some other role, their differing roles will bring with them different purposes, and these could color their sensemaking. As we showed in our study, this consideration percolated up in the findings.

The Take-Away

While there is value in all of the works cite above— their rosters of stakeholder types, and their commonsense mappings of stakeholder types to explanation requirements — there have been only limited attempts to investigate these matters empirically (e.g., Kaur, et al., 2020; Liao, Gruen and Miller, 2020). For example, the entries in Table 2 (above) for the Sensemaking Needs and Explanation Requirements were all reasonable, but nonetheless speculative. We therefore sought an empirical foundation for verification and expansion of this sort of analysis. The Stakeholder Playbook is a synthesis of our results, and we present that following a discussion of our Method and Results.

3. Method

Participants

Given the emphasis of XAI research on generating explanations for End-users, an attempt was made in our research to solicit the participation of individuals who represent diverse stakeholder groups, particularly in such roles as System Developer, Program Manager, and Development Team Leader. Participants were 18 professionals (16 males, two females) who had experience with AI and/ or autonomous systems. The group of participants included former military, civilian scientists working for the government, scientists working in the private sector, and scientists working as independent consultants. Participants were all either mid-career or senior professionals, and all had postgraduate degrees. Participants were recruited by soliciting individuals with appropriate experience and expertise, via relevant industry and DoD contacts of the research team, and were not paid to take part in the study.

Table 3 describes the kinds of AI/Autonomous systems with the participants had experience. As this Table shows, the pool of participants had experience with a diverse range of AI/autonomous systems.

Table 3. The types of AI/Autonomous systems with which each participant had experience.

ID	Previous/Current Roles	Types of AI Systems Experienced
1	Supports legal professionals and regulators	NLP systems for dictation; Spam filters
2	Supervisory role in human-systems integration and evaluation; Acquisition support	Tactical radios, Command & Control systems; Machine Learning systems
3	Contract formation, defense in litigation	Command & Control Systems

4	Electromagnetic Warfare; End-user; System evaluation; System evaluation; Capabilities development	Battle management systems; Tactical radios
5	Conflict analysis and resolution; Data science legal issues	Legal applications for client decision-making
6	Data analytics; Software development; End-User; Guidance to large organizations on conflict resolution	Predictive modeling systems for intelligence analysis
7	System integration, AI system policy; Manpower; Human-Machine Teaming	Image recognition; UAV control
8	Guidance for businesses to implement strategy and tactics.	Aircraft systems; Counterintelligence systems
9	Development team leader; Systems Analysis (human-automation systems); Applications (of AI)	Simulation-based training systems; Intelligence analysis; Decision aids
10	System Development; Modeling and Simulation Systems;	Predictive modeling at the organizational level
11	Intelligence analysis systems	Decision aids; Command & Control systems; Anomaly management
12	System Development; System Evaluation	Mission planning systems, Visualization systems; Planning optimization systems; Systems Evaluation
13	System Development; System Evaluation, Development Team Lead; Program Management; User Experience Evaluation	Network systems, Robotic Systems; NLP systems; Information Management Systems
14	Development team leader; Design based on human factors; User Experience analysis ; System Evaluation (Usability)	Command & Control Systems; Course of Action Analysis Systems
15	Development Team leader; System Acquisition; System Development (design)	Visualization systems, human-automation collaboration
16	Knowledge Management	Autonomous Systems; Command and Control Systems; Decision Support Systems; Systems Evaluation
17	Strategic planning; System Evaluation	Cyber Defense Systems; Intelligence Analysis Systems
18	Development Team Lead; Program Evaluation	Anomaly Detection Systems; Pattern Recognition Systems; NLP Systems; Machine Learning Systems;

Table 4 lists the participants' key demographics: Age, degrees, current and previous job description or title, current and previous self-identified role(s). As Table 4 also shows, the participants represent diverse roles and stakeholder groups. Most of the participants had experience in more than one role. We had a participant who had risen to the post of Development Team Leader and had been trained in cognitive science as well as computer science. Two participants had been trained in experimental psychology, but moved into applications, becoming Cognitive Systems Engineers and System Developers and one Participant had been trained in Industrial Systems Engineering but moved into Cognitive Systems Engineering and the role of Systems Developer.

One participant had a background in Human Factors of AI applications. Another participant had been trained in mathematics, but moved into system design. Four of our 18 participants self-described as End-users although their current primary role was not that of an End-user. Thus, occasional comments by a participant might be from their previous perspective as an End-user when their primary current role was, say, that of a Developer.

Table 4. The participants' key demographics.

ID	Age	Degrees	Previous/Current Job(s)	Previous/Current Roles
1	41	BA, Ph.D. Computer science	Professor of Computer Science	Supports legal professionals and regulators; Provides guidance on how the AI conducts decision analysis
2	46	Ph.D., Computational Modeling	Branch Chief; Leader of Field Study Teams; Development Team Leader	Supervisory role in human-systems integration and evaluation; Acquisition support
3	38	MA in procurement law; J.D.	Procurement Contracts law	Contract formation, Defense in litigation
4	47	Postgraduate	Warfare Center Director	Electromagnetic Warfare; End-user; System evaluation; System evaluation; Capabilities development; System Integration
5	36	Ph.D. in Law and Economics	Law Assistant Professor; Lab Director	Conflict analysis and resolution; Data science legal issues
6	30	MA in public policy, terrorism	Operations Research Analysis	Data analytics; Software development; End-user; System Evaluation; Guidance to large organizations on conflict resolution
7	46	Ph.D. Engineering	Assistant Director of AI Programs, End-user, Policy	System Integration; AI system policy; Manpower; Human-Machine Teaming
8	49	MA in Logistics, MA in Strategics	Independent Consultant; Formerly ranking officer in a security and intelligence organization	Guidance for businesses to implement strategy and tactics; End-user; System integrator (by default)
9	52	Ph.D., Experimental Psychology	Cognitive and Complex Systems Engineer	Development team leader; Systems Analysis (human-automation systems); Applications (of AI)
10	55	BA, Ph.D. in Electronics Engineering	Program Manager	System Development
11	60	Ph.D. Industrial Systems Engineering	Principle Systems Engineer Design Lead	System Development
12	68	Ph.D., Psychology	Consultant in Cognitive Systems Engineering	System Development; System Evaluation

13	52	Ph.D., Information Systems	Division Lead; Development Team Lead; Program Management	System Development; System Evaluation, Development Team Lead; Program Management; User Experience Evaluation
14	67	Ph.D., Psychology	Research Psychologist; Human Factors Scientist; Consultant	Development team leader; Design based on human factors; User Experience Analysis; System Evaluation (Usability)
15	64	Ph.D., Mathematics	Lead Scientist	Development Team leader; System Acquisition; System Development (design)
16	48	BA, MBA Engineering; International Relations	Program Management; Policy, Consultant	Knowledge Management
17	47	Ph.D. Cognitive-Experimental Psychology	Chief of Laboratory; Research Professor	Cyber Defense; Organizational Strategic Planning; System Evaluation
18	45	Ph.D. Cognitive-Experimental Psychology	VP of Business Strategy; Program Management	Development Team lead; Program Evaluation

Table 4 shows how most if the Participants have had multiple roles in evaluating, developing, procuring, or using AI systems. A Developer will need to understand conditions under which an AI is reliable, and a Legal Rights Advocate would need to know those conditions as well. But the two roles involve purposes and goals. The Developer needs to explore the reliability conditions in order to modify and refine the AI system. The Legal Advocate needs to be able to frame arguments about whether and when the AI can be trusted. End-users often are System Integrators since their work system involves many AI components which "talk to" one another. End-users sometimes play a key role in procurement decision-making.

This affirms the value of regarding stakeholder categories as conceptual clusters of convenience, but regarding *roles* as the critical determiners of how an individual will need to understand the AI system. This shift in thinking has implications for the practicalities of categorizing the responses collected in this study. A given participant's statement might express the explanation requirements of an End-user, but may itself be expressed from the perspective of a Developer. Indeed, our participants frequently referenced the explanation requirements of other stakeholders. Additionally, the explanation requirements of certain roles overlap with other roles. For example, End-users sometimes play a key role in procurement decision-making. As another example, Jurisprudence specialists are more like End-users than Developers or System Integrators, because (as one participant put it) they are "OK with some of the stuff being a black box". But they need to have appropriate trust and appropriate mistrust and be able to determine when they have entered a grey area. In short, the roles are not equivalent to professional identity, and the roles are fluid and sometimes people serve in several roles simultaneously.

Procedure

Participants were asked to participate in a brief interview concerning the explanation and understanding of AI systems. The Consent From expressed the topic this way:

> "It is sometimes said that Artificial Intelligence systems are "black boxes." One cannot see their internal workings and develop trust in them by understanding how they work. The general goal of this research is to adduce information about how people understand the AI systems that are used in their workplace. This information will enable us to tailor the explanations of how the AI works, depending on the individual's role, responsibilities and needs."

First there was a brief discussion of the ambiguity of what AI means to achieve some common ground. Next, there was a brief discussion of the various "stakeholders" (in addition to End-users) who need and benefit from explanations of how an AI system works. This discussion had the purpose of deciding which "hat" or "hats" the interviewee was most comfortable representing. Then, three demographic questions(age, educational background, and current job title or description) were asked.

Interviewees then engaged in a discussion based on the following questions:

1. What are your current responsibilities?
2. What Artificial Intelligence systems have been a part of projects on which you had responsibility?
3. How was it explained to you how those AI systems work?
4. What do you feel you need to know about an AI system in order to properly exercise your responsibilities?
5. Can you briefly describe any experiences you have had with AI systems where more knowledge would have helped?

As each interview proceeded, there were some slight role-dependent adjustment to the wording of some of the questions. For example, Jurisprudence professionals were asked *What legal issues regarding "Artificial Intelligence" systems are of concern to you?* Policy Makers were asked *What policy issues regarding "Artificial Intelligence" systems are of concern to you?* System Developers, System Integrators, and Program Managers were asked *What do you feel you need to know about an AI system in order to properly procure systems or manage system development?*

The interviews took between 15 and 50 minutes, averaging 23 minutes.

4. Results

The most informative way of conveying the results would be to provide extensive quotations from the interview transcripts For brevity's sake, however; we provide only representative examples.

Sensemaking Requirements and Challenges

For ease of exposition in laying out the results on Sensemaking Requirements, we clustered the participants as:

(1) Jurisprudence and Contracting. These roles tended to focus on evaluating and anticipating the impact of a complete working system, and how it fits into the organization's processes.
(2) System Development, Evaluation, and Integration. These roles tended to focus on wanting causal accounts of how systems worked from input to output, in the relevant contexts.
and
(3) End user, Developer, Integrator. These roles tended to focus on how the tools can be used to assist in the work they are trying to achieve.

Table 5. Sensemaking Requirements and Challenges as a function of Role. The italicized text is quotations from the interviews.

SENSEMAKING REQUIREMENTS AND CHALLENGES
Jurisprudence, Contracting (P1, 2, 3, 5)
These roles tended to focus on evaluating and anticipating the impact of a complete working system, and how it fits into the organization's processes. *If you can't look under the hood, you can't evaluate them. How transparent is it?* *What data is being fed into it and how does it work?* *The most important thing I need to know about an AI system is the features that were relied upon.* *Where the AI makes decisions or judgments, are there biases, if so, what are they, and how can they be corrected or mitigated? From what do the biases arise? The input data?* *Legal/contracting needs to know "how it works" but the challenge here is lack of computer science knowledge, lack of knowledge about AI/ML as that area is just ramping up.* *Rather than interfacing through the salesmen, I would go and sit next to the developer and work with them. And they were very helpful.* *The mathematicians have to explain to me what they do. I bug them a lot to give me non-technical language on how it works.*
System development, evaluation, integration (P4, 6, 10, 11, 12, 14, 17)
These roles tended to focus on wanting causal accounts of how systems worked from input to output, in the relevant contexts. *I would like to know [chuckles] exactly how the algorithms work, what information they are using, how they manipulate it, and how do they get the results they are getting. I have to see all the way into the black box, absolutely.* *I need use cases that are representative of the implementation contexts, user groups for the test bed, desired development tools.* *I need explanation about the data that were used to train the model, the model's fitness for the data that is used in the production environment. I need to understand the fitness between what the model was trained on and what you are analyzing today. I need explanations of the confidence or applicability of the algorithm for the particular question that is being asked. For example, in data analysis, a search yields some results. You do not know whether the model is well fit for the question you asked of the search tool.*

I have to know how it works in order to get feedback from the users. If I don't know how it works I do not know how to frame the usability test to collect the data in order to improve the system as we are developing it.
End-user (also developer, integrator) P6, P11
These roles tended to focus on how the tools can be used to assist in the work they are trying to achieve. *I test the tools to find out how they work and how they apply. Part of it is trying to find out how they work, part of it is trying to find out how they apply.* *What I have found is it is usually either they are going to say the system just uses machine learning, or they say "We take these variables and we give you the answer."* *So, it is either very low-level or it is very specific and over my head because I did not develop the tool and I also am not an expert in the specific modeling that they are talking about.* *And so, rarely have I seen... it is not common for someone to walk you through how it works.* *A lot of times I'll poke around with it about how it works, see if I can find out... the things that make it tick.* *If I do not understand what it is I'll not use it.* *It kind of ends up me having to do it on my own, or poking through it on my own.* *Rather than interfacing through the salesmen, I would go and sit next to the developer and work with them. And they were very helpful. Most people want to answer your questions, especially someone who develops it.*

Some participants commented that they have received explanations that are good. In contrast, one participant (P2) asserted that they simply do not need explanations. This came as a surprise, insofar as it suggests not all roles actually want or need explanations. Five participants asserted that the explanations they are provided were always or often insufficient, sometimes too low a level, sometimes too detailed (P6, 9, 11, 13, 15). Four participants asserted that they (or other stakeholders) do *not* need to be able to drill down into the technical details of how the system works (P 2, 8, 12, 18).

Information Access as an Explanation Requirement

Self-explanation is often triggered by the inadequacies of explanations that are provided.

The superficial answer--it takes into account these variables, and chugga-chugga, and we're not going to explain that, and then the other [explanations] go way down

> *into the weeds and not necessarily for you a direct path to understanding how it works. And so you just dive in and use it, right? Honestly, for me I'm just very interested in about how a lot of these tools work. So I don't like using something that I can't explain. So I am giving an explanation to, say, a commander. "I do not really know how this answer came out, but this is the answer." So a lot of times I'll either poke around with it about how it works* (P6).

Self-explanation is both important and pervasive. In general, people engage in an exploratory effort in order to self-explain the AI because inadequate explanatory information is provided or is available. And when it is, it is either superficial or so detailed that it is not useful. Thus, the challenge of self-explaining often becomes a matter of access to the right people who are in possession of the right information.

The motivation to self-explain is often manifest as active reach-out to the Developers:

> *I sought them face-to-face, verbal. For the most part, they were satisfying. Global explanations. And you had to seek them out; they were not provided anywhere. I found help via professional networks* (P17).

The motivation to self-explain is manifest when the available local resources are inadequate:

> *I've gone to YouTube channels and had some random person from across the world explain some really complicated topic explanations and because their entire goal is to explain the modeling process, I understand it better than anyone else has ever explained it before. There's little things like that. You can see what other people are doing to explain it. It's like, "Oh, wow, I can learn this better from YouTube than from anybody else"* (P6).

A number of our participants commented that the establishment of a sufficient and satisfying understanding can only be achieved via discussions with the system Developers:

> *[I'm] glad to hear that other people have these experiences. I'll look around the room and think, "Am I the only one that's not getting this? Am I the only one who is confused?" One of the things that has been at least helpful for me is, a couple of things we have worked with the development people, and we were closely tied with the actual developers* (P6).

A number of participants commented that their effort at self-explanation often serves as useful feedback to Developers:

> *Additional explanations, beyond the demos, provided by the developers. Wider range of examples to show the results based on different input conditions. Putting it through the rigors of the wider range of potential situations that might encounter. We often ended up doing it for the developers through our simulated exercise. Even repeated scenarios are not exactly the same every time* (P11).

Because of its importance and salience to our participants, the Stakeholder Playbook presented below) includes "Access Requirements" and well as "Explanation Requirements."

Trust and Reliance

The participants were not explicitly asked about trust and reliance, but all but one of the them did mention trust issues, which is not surprising given that trust issues are closely related to the challenges of explanation.

Trust takes time:

> *I never totally understood how it worked, but I worked with it enough to trust it* (P7).

But trust is sometimes a "default":

> *A lot of it you take on face value because you are not going to become an expert End-user in all the beeps and squeaks that go into it. You can't have users dive into everything. They just want to know that it is going to do what they are asking it to do, and the more they see that, the more they develop trust* (P7).

> *It does things sometimes you don't know* (P4).

On the other hand, trust is very often tentative or skeptical, and different End-users have different default stances along a spectrum of trust-distrust when using a new system:

> *People within our own organization say "Hey, this is really awesome." But then I would say "Yeah, but how much work am I going to put into that to get this tool to answer my question?"* (P6).

> *You have to overcome the initial trust hurdle early on, overcome it with training* (P7).

> *There are some scenarios where you know it was going to act but not which way it was going to act* (P7).

> *The challenge was figuring out what the system was going to do. The operator had to figure that out, to know what it was going to do (because you were not going to be able to affect it)* (P7).

> *Where you feel you have an input, and it does not give you back what you believe, it confuses more than helps* (P8).

> *I learned early on that all of the systems were brittle, even systems that were considered a success--some evidence that they were resilient, adaptable. But I knew they were brittle but did not really know why* (P9).

> *It comes down to how much faith you have in the individual [who is providing the AI system]. Is the decision the machine makes the same as I'd make, or does the AI think better because it is aware of other things? You need to trust the brain behind it, and their understanding of how it works, what its rules are* (P16).

> *I am willing to start out trusting, it, until it proves itself to not be reliable. Then it becomes trust but verify. Trust and Reliability feed each other. You start off trusting and watch reliability over time. Or you can start relying on it. Is a matter of loss of trust. But if it is life or death, you do not start with trust, but if it is a simple decision like data analysis, you can start with trust* (P16).

And trust is not just in the AI; it extends to other people and to entire organizations:

> *Contractors promise us the world. And then it turns out, they do not necessarily have that. So some of the tools that we would work with didn't really end up doing anything* (P6).

> *People within our own organization say "Hey, this is really awesome." But then I would say "Yeah, but how much work am I going to put into that to get this tool to answer my question?"* (P6).

> *Companies were trying to push AI but no one wanted to hear about the limitations, constraints, and brittlenesses. The companies try to push things further, but no one wanted to talk about the warts or the limitations* (P11).

And trust is not just in the AI, it is in the data:

> *Need to know whether there is bias, and then want those biases are. It is necessary to know about the data that go into the AI system* (P3).

> *You need to know what data it evaluates, in order to quantify its uncertainty. Data are not free. Data for AI ingestion is often not properly curated, tagged, etc. That is not cheap, it is certainly not free* (P7).

> *I'd seen that multiple times, and know I can't rely on that data but this other part of the data is valid.*

Trust extends across stakeholder groups.

> Trust in the vendors trumps the need to poke around. Conversely, mistrust in the vendors (because of their over-promising) trumps the AI's reliability and trustworthiness. Vendors need to instill confidence in the prospective adopters and users, even if the vendor cannot explain the "under the hood."

Finally, training on "edge cases" is crucial for the establishment of trust:

> *What you are trying to find is where those edge cases where it hasn't been fully vetted... it's reaching the limits of its data that it using to inform its decisions. That's what everyone is going to be concerned about. If it lives in the heart of the black box the whole time, then you will have adequate trust it will perform the way it is expected to, but for your policy makers and your legal, they will certainly be interested in what those edge cases are. And the users will be, depending on how quickly they reach those edge cases in one of the scenarios* (P7).

Looking across the interview data, and our multiple ways of "slicing" the data, a number of findings stood out as surprises. These are elaborated below.

5. Surprises

It has been assumed in many XAI activities that End-users and other stakeholders all need explanations. Our results reveal a far more complex and rather subtle picture. What many of these surprises show is that choice of algorithm is a small part of what is needed to support a given stakeholder. There are explanation requirements that go beyond the need to know how a particular algorithm works.

Explanations That Are Provided Are Rarely in a Goldilocks Zone

When an explanation is desired, the explanations that are provided are generally regarded as inadequate. Five participants asserted that the explanations they are provided were always at either too low a level or too detailed (P6, 9, 11, 13, 15). Either way, the stakeholder needs to reach out for more information. A number of our participants referred to their active reach-out to other people and other sources (e.g., Developers, other End-users) and other sources (social networks, YouTube) to enrich their understanding of AI systems (P4, 6, 7, 9, 10-17). XAI research has involved the creation of explanations that take a number of different forms and express different kinds of content. Formats include diagrams, heat maps, matrices of feature weights, logic trees, etc. Our participants did not volunteer any opinions about such formats, except for the occasional reference to the need to "visualize" data.. What they did refer to, and often, was dialog in which they sought out and received explanations from other people.

For individuals who are not particularly computer savvy, global explanations can take the form of reductive but clear analogies. The value of explanation-by-analogy is crucial in scientific reasoning and problem solving. We were surprised that analogy only appeared once in our interviews, when a participant referred to a "sitting kids in a school bus" analogy to describe how ML systems work. Explanation by analogy has been underplayed in the XAI work.

"Global vs. Local" is Not Clear-Cut

The distinction between global explanation (How does it work?) and local explanation (Why did it make this particular decision?) has been a key consideration in the literatures on explanation and in the work on Explainable AI (see Miller, 2017). One of our participants (P17) asserted that they only need global explanations (high-level accounts of how the system works, in contrast to explanations about why a certain decision was made). This global focus appears widespread, as most comments by participants relate a desire to know *how the system works,* rather than wanting to know *why it made a particular decision*. This is in stark contrast to the vast majority of research in XAI, which focuses on local explanations and justifications of decisions or actions of an AI system.

Previous research has shown that global explanations are often illustrated by specific cases, and that local explanations contain hints that contribute to global understanding (Klein, Hoffman and Mueller, 2022). In other words, people benefit from having global and local explanations that are integrated. Participants in the present study made more reference to global explanation than to local explanation, yet the majority of XAI systems of which we are aware are focused solely on the delivery of local explanations.

Our participants' frequent reference to the need to see "edge" cases underscores our finding that explanatory value derives primarily from material that blurs the global-local distinction.

Comments from our participants underscored this finding. For example, one participant (a Developer), said he benefitted from having global and local explanations that are integrated.

> *Going from the specific allows me to go to the general. Enough specific examples allow me to accept that there is a general explanation. (P10).*

Spoon-Feeding versus Exploration: Explanation versus Understanding

We asked our participants *How was it explained to you how AI systems work?* and *Can you briefly describe any experiences you have had with AI systems where more knowledge would have helped?* Answers to these questions revealed a complex and subtle picture.

One participant (P2) asserted that they do not need explanations. This came as a surprise, insofar as it suggests not all roles actually want or need explanations. Four participants asserted that they (or other stakeholders) do *not* need to be able to drill down into the technical details of how the system works (P2, 8, 12, 18). As we planned the interviewing we expected that all of our participants would say they want and need explanations, and better explanations than the ones they are typically provided. This expectation on our part was largely due to the fact that the "explanations" we saw being generated by XAI systems seemed exclusively local and techno-centric (e.g., heat maps, matrices of feature weights, etc.). When we asked our participants, *Can you briefly describe any experiences you have had with AI systems development where more knowledge would have helped?,* three participants said *All of them* (P9, 11, 15); two participants said *Yes* (P10, 19), another participant said *Absolutely* (P13), and another Participant said *Yes, always actually* (P15). These responses were terse, immediate, and strident. Other responses were a less terse and more informative:

> *What I have found is it is usually either they are going to say the system just uses machine learning, or they say "We take these variables and we give you the answer"* (P6).

> *A lot of the explanations I have heard, always stop short. I need to know exactly what the AI is thinking when it gets into a scenario and something happens* (P16).

Another participant affirmed the question but was less conclusive about it:

> *[There] probably were times when I did have to deal with an AI and did not know a lot that was going on* (P8).

What was surprising was that we did not get strident affirmations of our probe question from all 18 of our participants. In response to our probe questions, one participant responded by saying: *Nothing immediately comes to mind. It develops as you drill down into the details* (P7); and another participant denied the need for a deep explanation: *I don't care much about the details of the algorithms. No situation where I felt a need to understand the algorithm in detail* (P12).

Indeed, more frequent than affirmations were assertions about the sensemaking needs of stakeholders *other* than themselves (P1-7, 10, 12, 13, 14, 15, 18):

> *Now that I've worked with but I know lots of law firms who are working on products and is not obvious to the outside world how the AI works*) (P1).

> *The practitioners try to get into how the AI/ML system works* (P2).

> *This is so funny*. [Laughs] *Yes. I have seen, for example even with networking systems. If the users understand how they work, it is easier to do troubleshooting and preventative maintenance checks. If they don't understand it, they are not going to be able to use it* (P14).

And muddying the waters yet further, four participants denied that one or another stakeholder needs an explanation (P2, 5, 8, 18).

In general, all of the participants referred to the need for knowledge, but this was not expressed as a need to be spoon-fed more or better explanations. One participant referred to explanation but seemed to be talking about understanding:

> *The universe in 30 seconds. [You] need explanation about the data that was used to train the model, the model's fitness for the data that is used in the production environment. Need to understand the fitness between what the model was trained on and what you are analyzing today. Need explanations of the confidence or applicability of the algorithm for the particular question that is being asked. Need automated support for visualizing the data, to understand how the data you are giving the tool might be inadequate* (P17).

And many of our participants said they would obtain that understanding by exploration, by reaching out, and by self-training:

> [I] *test the tools to figure out how they work* (P6).

> *When I needed to I could go to them* [engineers] *for explanations* (P4).

> *It comes down to how much faith you have in the individual* (P16).

> *[I got] explanations in the past. I sought them. Face-to-face, verbal. For the most part, satisfying. Global. And you had to seek them out; they were not provided anywhere. Found help via professional networks* (P17).

> *I have to know how it works in order to get feedback from the operators. If I don't know how it works I do not know how to frame the usability test to collect the data in order to improve the system as we are developing it* (P17).

> *The training is baked into the system; training is embedded in the software. So if someone does not know how to use it, they need to be able to train themselves on how to use the tool* (P2).

So, what to conclude? Do stakeholders want explanations? Do they want more explanations? Do they want better explanations? The answer is Yes and No. Because the explanations they are spoon-fed are often deficient and insufficient, they do not want to be spoon-fed more such explanations. What they want is a richer understanding. Understanding is the goal of the sensemaking process, not the comprehension of a piece of text or the perception of a heat map. An explanation should have explanatory value, which can contribute to sensemaking. But there is more to the achievement of understanding than the comprehension of a spoon-fed explanation, whether generated by a machine or by another human.

What stakeholders want is empowerment. End-users and other stakeholders need sufficient global and local (case exemplar) information to enable them to self-explain by discerning where and how to explore and where and how to reach-out to others. This key finding is manifest in the details in the Stakeholder Playbook.

Stakeholders Can Be Quite Interested In "Deep Dives"

It has been assumed in XAI activities that End-users and other stakeholders do not want, and are not prepared to understand highly detailed or technical explanations of what goes on "under the hood." Our findings show that this is not always the case. Although End-users and User Team Leaders do not always have a sufficient knowledge of computing concepts, our findings show that stakeholders are often sufficiently versed in computer science to enable them to do deep dives inside the ML system (e.g., why it makes certain kinds of errors). Thus, some stakeholders sometimes want detailed technical explanations (P1, 4, 6, 12, 14, 15) and are often sufficiently versed in computer science to enable them to do deep dives (e.g., why it makes certain kinds of errors). They recognize the value in their being able to do this.

Our findings show that stakeholders' cross-disciplinary skill is perhaps more common than might be supposed. Yet, there are certainly circumstances in which a stakeholder has little understanding of AI and is frustrated because they might not know exactly how the AI is making the decisions (e.g., P3).

Sensemaking By Exploration Is Of Greater Interest Than Prepared Explanations

Stakeholders in all roles actively seek out satisfying explanations. The End-user has to engage in an exploratory effort in order to self-explain the AI, because inadequate information is provided or not enough information is available. And what is provided is either very superficial or so detailed that it is not useful. Half of our participants asserted that they (or other stakeholders) are self-motivated to actively develop good explanations of "how it works." (3, 4, 6, 7, 12, 13, 15, 17). Eleven of the 18 participants asserted that there are circumstances in which they (or other stakeholders) need to be able to actively seek and then drill down into the technical details of how the system works (4, 5, 6, 10, 12, 13, 14, 15, 17, 18).

It is widely recognized that contrastive explanations are valuable: "If X had been different, what would the AI have done?" or, "Why did the AI decide A and not B?" (Miller, 2017). The closest any of our participants came to this was in two kinds of statements: (1) statements about the need to understand how the AI would perform if the input data were of questionable quality, and (2) statements about the need to understand how the AI performs when dealing with "edge cases." The participants take it for granted that the AI would do something differently if the data were different. Contrastive explanation did not take a logical form for our participants, but was of

an exploratory nature. A number of participants commented about how they preferred to manipulate ("poke around") and explore the AI system behavior under different scenarios, to "get a feel for it." Stakeholders want to be provided with more examples of the AI encountering different situations. End-users would benefit from local explanations that are exploratory: The visualization of tradeoffs (e.g., in a scheduling algorithm) would support appropriate reliance and the capacity to anticipate when anomalous events occur and the recommendation may be misguided. Global explanations are not just for the purpose of understanding--they enable the search for other necessary information and resources.

Explanations Are Often Needed For Presentation To Other People

As we just mentioned, Developers do not always want explanations for themselves, but to help End-users create better understanding. Evaluators engage in self-explanation and exploration, but not for themselves, but to help Developers create better explanations for the End-users. Explanations provided to leaders (e.g., a commander) have to focus on the rationale for the answer and why the answer makes sense, not on how the AI works.

> *We need to be able to explain the user or his higher. At some level we need to explain to whoever is going to take the system over. "Here's why we did it this way" to get them to be able to think about the trade-offs we are thinking about* (P17).

> *I need to know enough to be sure I understand when I am talking to other non-technical stakeholders, what capabilities are afforded by the systems they are funding us to create. If not, I would have a hard time describing the relative pros and cons between one approach and another, or describing the success of the approach we are using* (P18).

> *So I am giving an explanation to, say, a commander. "I do not really know how this answer came out, but this is the answer." So a lot of times I'll poke around with it about how it works* (P6).

Stakeholders Are As Likely To Need To Know About The Data As They Are to Need to Know About The AI System That Processes The Data

Rather than expressing a desire to understand the inner workings of the AI itself, more common was participants' expression of a need to know about the input data or the data features the AI processes, or a need to see to see the input-output relations in additional scenario-demonstrations (P1, 3, 4, 5, 7, 8, 11, 14, 16, 17, 18). Understanding the data the AI system uses would be more helpful than poking under the hood to examine the innards of the system. They wanted to know what data were used to train the AI/ML. They want to know about any system biases. They want to know what data were used for a specific project or decision, and they want assurance that there is a match between the data inputs and the situation—if the AI/ML has been trained on or is using the wrong data, the outputs can't be trusted.

Stakeholders Need To Know About How The XAI Interfaces With Other Systems

Stakeholders expressed interest in explanations that are not about 'the AI' but rather about the overall system architecture and business logic of the system. For example, what are the plug ins? How is it interfacing with other systems? How is it connecting the different units within an

organization? How is it synthesizing their data? What is the interplay of the components? What is the AI/ML accomplishing? What is the Cost/Benefit tradeoff of using the system?

Explaining Is Not A "One-Off"

An additional limitation of prepared explanations derives from the assumption that a global explanation is provided once, likely during training or at the beginning of operational experience. However, the AI is always dependent upon or integrated with other systems. Explanations are often context-bound, trust is always tentative and explanation is not a process that terminates. Explanations are never one-offs, but must be persistent and engaged in frequently, especially for AI/ML systems that learn and change (improve*).*

> *Processing tools are very complicated but highly reliant on data streams and samples, need to understand what you are feeding it. For example, data collected from sensors undergoes initial processing that may result in gaps, may be collecting the wrong kinds of data, data may be skewed and misleads the AI. I need to understand what the data look like that are going into the model. I might be able to tune the sensors, or put context-specific labels on the data* (P17).

The Word on the Street: We Set A "High Bar" For XAI Systems

Participants referred to stringent criteria. Incremental gains just aren't worth it. The AI system has to be a game-changer. One participant asserted that if a novice user cannot perform 85% of the key tasks on the first test, the system will not be adopted. A number of participant comments were candid expressions of the actual sentiments of operators:

> *Can a user, without training, figure out how to use the system within ten minutes? If they fail at that, they don't use it* (P2).

> *I don't like using something that I can't explain. In some few cases I did get at least enough understanding that I felt comfortable for explaining to someone else. I give them an answer and I say, "I used this tool and got this answer. I'm not the expert obviously, but here's the general idea of what happened and why it makes sense."* (P6).

> *The tech has to be a game changer, not just an incremental improvement, it has to be a substantial improvement in performance because it is so hard to introduce and maintain new technology. The tools have to be a leap-ahead* (P5).

Trainers Need To Be Able To Train End-users On "How It Fails" And "How It Misleads" (Limitations And Weaknesses)

Trainers—and other Stakeholders—need to have an understanding that is sufficient to enable them to explain the AI system not to themselves, but to other people. Trainers need access to a rich corpus of cases that are representative of implementation contexts, of course. But they also must be able to train End-users to engage in maintenance and troubleshooting activities. Trainers must be able to train End-users to sense how quickly they will enter a grey area in various

scenarios. The XAI needs to do more than explain the "why" of particular decisions: It needs to be able to give the user advance knowledge of when the work system is approaching an edge case.

Stakeholders Need To Understand The Design Rationale

Stakeholders need to know the rationale for the answer and why the answer makes sense (P6, 14). This is expressed in terms of the domain (concepts, principles, causes, etc.) and not just in terms of how the AI works. They need to be able to explain the design rationale to Vendors.

Operators are Often Left Adrift

Our participants expressed the sentiment that even if a system is thought to be useful in general, the need to integrate a system into their organization's processes and mission is ignored and the integration burden falls on the operators.

> *The search for tools that really help with the job takes priority. Only after that is there a matter of exploration and self-explanation. Part of it is trying to find out how they work, part of it is trying to find out how they apply* (P6).

> *People would come to me with tools, even people within our own organization and say "Hey, this is really awesome." I would say "Yeah, but how much work am I really going to put into that to really get this tool working to answer your question?"* (P6).

> *I have to jump to see what if any of the tools might apply to any or all our organization's problem sets* (P6).

The System Designer Needs To Know About The End-user

It is crucial for Developers to understand the legacy work, how the End-users do what they do using their legacy system. It is important for the Designer to get End-user reactions to the AI/ML system. End-users can help Developers re-create the conditions that led to confusion and problems.

6. Reconciling the Paradoxes

Some stakeholders want explanations, some do not. Some want detailed technical explanations, some do not. Some expressed a need to explore, some did not. Some stakeholders say they only need a global understanding, but some say they do need to look under the hood.

How can such contradictory findings be reconciled? A given individual may not need an explanation, either global or local, depending on their style and circumstance (e.g., they can rely on trusted Developers). But individuals in all roles do want and need satisfying understanding of something, either the AI or the date that are fed to it, at least some of the time. Their expression of a need for explanation (or self-explanation) is often subtle and indirect. But *not everyone needs explanations*, and *explanations are not needed all the time*. In fact, only half of our participants expressed an interest in receiving explanations. The others were either indifferent or were explicitly disinterested in receiving explanations. These results paint a much more chaotic picture

about stakeholders than much of the previous theoretical and taxonomic research we reviewed earlier.

So, can these contradictions be resolved? We believe the answer is "yes," and it can be done by acknowledging that different stakeholders have different capabilities, different sensemaking requirements, and different immediate goals. Different cognitive styles also play a role, as suggested by participant comments to the effect that they preferred to dive in and play with the system, rather than getting an explanation of how it works. These factors combine to define what, for each individual, constitutes satisfactory and actionable understanding.

7. The Stakeholder Playbook

Although there are somewhat differing explanation requirements for different stakeholder roles, there are also some similarities across stakeholder groups. For each stakeholder group, information is presented about the requirements for explanations that we derived from interview notes and analysis of the different stakeholder roles. Note that the Playbook itself is agnostic with regard to the source of the required information. The source might be the AI/XAI system but it might be some other person. For each stakeholder role, information is presented about the requirements for access to key individuals who can provide clarifications or additional explanations. For some stakeholder roles, the Playbook also includes "cautions": Things that the explainer (or the XAI system) needs to be cautious about. The goal of this playbook is twofold: to help an XAI developer understand concrete examples of potential stakeholders, and to provide initial guidance and cautions for explanations that support the different stakeholders. Note that we intentionally do not provide specific recommendations of explanation algorithm or explanation types for different stakeholders, because many different approaches might fit the needs of distinct user roles. Moreover, a particular algorithmic explanation is often only one small part of the explanation needs of a user.

JURISPRUDENCE
Explanation Requirement: Analysis of system biases, assumptions, and bounding conditions.
Explanation Requirement: Description of the features upon which the system relies.
Access Requirement: Access to succinct background information on computer science and the pertinent AI technology.
Access Requirement: To the system development team— trusted software engineers, mathematicians.
Access Requirement: To experienced and trusted domain practitioners.
CONTRACTING; PROCUREMENT

Explanation Requirement: Global explanation of "how it works" and how the data are processed. Explanation Requirement: Global explanation of architecture and functionality (how the data are processed. Explanation Requirement: Analysis of system biases, assumptions, and bounding conditions. Explanation Requirement: Description and explanation of the data that were used to train the model. Explanation Requirement: Analysis of the model's fitness for the data that are used in the operational environment. Explanation Requirement: Analysis of the confidence or applicability of the system for the particular questions that are being asked. Explanation Requirement: Assurance of data quality and curation. Access Requirement: Trusted software engineers and domain practitioners. Access Requirement: Leads with technical background need access to explanations of technical details. Access Requirement: Access to trusted vendors, who do not "dumb things down."
PROGRAM MANAGER; DEVELOPMENT TEAM LEAD; DEVELOPERS
Explanation Requirement: Global explanations of "how it works." Explanation Requirement: Description of the data that the AI ingests; assumptions about the data. Explanation Requirement: Analysis of how the system will be integrated with other systems in the broader work system. Explanation Requirement: Analysis of system strengths, weaknesses, and bounding conditions (system assumptions). Explanation Requirement: Analysis of how the system will be integrated with other systems in the broader work system. Access Requirement: Preparation in computer science and AI programming — practice as well as training. Access Requirement: Access to trusted software engineers, mathematicians. Access Requirement: Access to experienced and trusted domain practitioners. Access Requirement: Group discussions among developers and management. Access Requirement: Trusted vendors who do not "dumb things down." Access Requirement: Corpus of use cases that that are representative of the implementation contexts. Access Requirement: Opportunity to explore the system by working between specific examples and global information; manipulate the inputs and see the outputs. Caution: Some Program Mangers and Development Team Leads will need to know the details of the system processes and algorithms.

SYSTEM INTEGRATOR
Explanation Requirement: Explanation at the detailed technical level; Needs to be able to "look under the hood." Access Requirement: To trusted software developers.
TRAINER
Explanation Requirement: Needs to be able to achieve an understanding that is sufficient for them to be able to explain t\the system to trainees. Explanation Requirement: Needs to be able to achieve an understanding that is sufficient to allow them to help users understand the edge cases, and when a situation is approaching an edge. Access Requirement: Rich corpus of edge cases.
SYSTEM EVALUATOR
Explanation Requirement: Explanation of the inputs, outputs and their relations. Explanation Requirement: Information of how the system manages the trade-offs in operational conditions. Explanation Requirement: Information supporting the design of usability and performance tests. Explanation Requirement: Operational definitions of the proposed "metrics" (measures) to be used in performance assessment. Access Requirement: Feedback from prospective users; Access to an established network of experienced users to support self-explanation. Access Requirement: The system developers when the system does something bizarre or unexpected. Caution: Not all evaluators need to understand the technical detail (e.g., algorithms).
POLICY MAKER
Explanation Requirement: Global explanation that is satisfying and consistent with how the system actually works. Explanation Requirement: Descriptions of system biases, assumptions, and bounding conditions. Explanation Requirement: Descriptions of system limitations and weaknesses. Explanation Requirement: Demonstrations that include edge case scenarios. Explanation Requirement: Need to know about how the system being evaluated was developed. Access Requirement: Demonstrations that cover a range of examples to show the results based on different input conditions. Access Requirement: A trusted system designer. Access Requirement: The system developers when the system does something bizarre or unexpected. Access Requirement: An established network of experienced users to support self-explanation.
END-USER; ADOPTER

Explanation Requirement: Global and local explanations that are satisfying and consistent with how the system actually works.

Explanation Requirement: Explanation of data inputs and how the system processes the data.

Explanation Requirement: Results of a cost-benefit analysis of different tools, with respect to the user's goals and responsibilities.

Explanation Requirement: Ability to explore the system behavior by "poking around."

Explanation Requirement: Explanations need to strike a balance between superficiality and technicality.

Explanation Requirement: Explanations that support troubleshooting and system maintenance.

Explanation Requirement: Intuitive displays for visualizing the data, to understand whether the data might be inadequate.

Access Requirement: An established network to support knowledge sharing and exploration.

Access Requirement: To the system development team— trusted developers and software engineers.

Caution: End-users sometimes do not care or don't need to know "how it works."

Caution: End-users often desire better explanations than the ones that are provided.

Caution: Continuing explanation is required as the input data, the work system context, or the operational environment change.

8. Conclusions

The Playbook presented here is a first pass. We look forward to further empirical efforts to collect more data such as we have, to elaborate on the requirements and to include other roles. Based on our sample, we conclude that the standard XAI tactic of providing explanations to stakeholders is not seen as valuable by a majority. Stakeholders prefer tactics other than being the passive recipients of explanatory materials.

In this spirit, the Stakeholder Playbook is intended to expand the horizons for XAI systems. By cutting loose from the assumption that explanations are "provided" and empowering the stakeholder's sensemaking, it must be possible to create better XAI systems.

Rethinking the Stakeholder Concept

Since most of our participants served or had served in multiple roles, assigning participants to role categories required multiple tabulations, depending on the participant's perspective when making particular comments. For example, four of the 18 participants self-described as End-users although their current primary role was that of a Developer. Thus, occasional comments by a participant might be from their perspective as a former or sometime End-user when their primary current role was, say, that of a Developer. This finding affirms our concern that the concept of a "stakeholder" may be misleading and even perhaps counter-productive. As this research topic moves forward, it may be more useful to talk about stakeholder roles than to pretend that people inhabit each roles exclusively.

Human-Centered Approach for Design and Evaluation

While concepts of human-centered computing and human-system integration have become a part of the Research and Development culture, human factors analysis and human-system integration are often tacked on as isolated, discipline-based practices. Procurement considers the design of technologies, but the technologies are part of a cognitive work system. Thus, a deep and thorough understanding of the legacy work is crucial. One of our participants commented: *The integrator needs to lead people who go into the field and see how users use network the technology*. Yet, vendors are not always required to study the legacy work and work context prior to designing the tools. How will the insertion of the technology change the cognitive work? What new forms of error might it trigger?

It is generally assumed that the main goal for XAI is to promote better performance by the human at the particular task that the AI is designed to accomplish, or help the human accomplish. But explanations need to support the process of human-machine integration as a contextualized work system in which the human and the machine are interdependent. The findings presented here call out the importance of a Human-Centered approach for design and evaluation.

> *AI/ML is early in the Tech maturation pipeline so, it is not in the systems that are being fielded and tested. But it's gonna be a big problem, how you assess this technology* (P2).

> *The challenge is how do you get AI/ML into the requirements documents; how do you design AI technology well for the user and get it into the requirement document* (P2).

> *We have limited time to evaluate all the time. We have novices who have never used the legacy system, but [on the new system] outperform the experts on the legacy system. They do not have the negative transfer and all that baggage. Experts on legacy systems are often slower on the new system because they are trying to figure things out, or it is not where they expected* (P14).

This is not just technology assessment, it is a human-system integration (HSI) issue. It is necessary to find the gaps in the user's needs, and understanding the methods to address those via human-system integration analysis. The Integrator needs to understand AI and ML systems from the standpoint of the work they are to help with. Participants called for training in the Acquisition community about usability and assessment methodologies. The End-user, especially the subject matter expert, needs to be able to do what the Integrator does. Explanation has to be about the development process and the design rationale, and not just about the AI system. The Integrator has to mentor the user-experts with regard to what the Human Factors person is doing—the explanation is of the process and rationale for the HSI activities. The User-expert needs to be able to do what the HSI person does.

Participants called for a procurement requirement to conduct systems integration.

> *For me, the overarching questions we should always be asking are: How do I use this? How will it help the task that has to be accomplished? How will it help the human accomplish that task? When we put AI in just because we can, we miss the overall objective. The objective is the task goal, why is it hard for humans to do, how will the AI make it better, and then based on that how will the human use the AI easier for the human do to. It may evolve the task or the task may become slightly different. But that human-centered*

> *or task-oriented design is at the center of designing any AI application is critical. This is often missed* (P7).
>
> *If it's really AI, it is supposed to be operating more in the complex cognitive realm. And then my skepticism kicks in, that the AI can really keep up and be useful to people. I would want to know from there, how much opportunity is there to allow the users finish the design* (P9).
>
> *I need to know all the things that the operators need to know. Oh, yes. Absolutely. I spend a lot of time understanding what they do, and I don't have their training. I get SMEs to work with me. And I mentor them to do what I am doing, So that they can do my job like I do my job, but they help me all along the way to understand everything and the "why" behind it* (P14).

The current requirements in the procurement process were not designed with AI systems in mind, and they seem inadequate to the task. How can AI requirements (including explanation requirements) get into the procurement process? How will AI/ML will require a change to the current "human readiness levels" scheme? How can we create procurement standards or "best practice" guidance for evaluating AI systems?

References

Al-Abdulakarim, L., Atkinson, K., Bench-Capon, T., Whittle, S., Williams, R., & Wolfenden, C. (2019). Noise induced hearing loss: Building an application using the ANGELIC methodology. *Argument & Computation, 10*, 5-22.

Al-Abdulakarim, L., Atkinson, K., & Bench-Capon, T. (2016). *Artificial Intelligence and Law*, *24,* 1-49.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115.

Arya, V., Bellamy, R.K.E. (and 18 others). One explanation does not fit all: A toolkit and Taxonomy of AI explainability concepts. [arXiv: 1909.03012.v2

Atkinson, K., Bench-Capon, T., & Bollegala, D. (2020). Explanation in AI and law: Past, present and future. *Artificial Intelligence*, 103387.

Dahan, S. (2020). "AU-powered trademark dispute resolution." Report to the European Union Intellectual Property Office (EUIPO).
[https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3786069]

Doshi-Velez, F., & Kim, B. (2017a). Towards a rigorous science of interpretable machine learning [arXiv:1702.08608v2].

Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M., & Hussmann, H. (2018, March). Bringing transparency design into practice. In *23rd international conference on intelligent user interfaces* (pp. 211-223).

European Union Commission (2016). "General Data Protection Regulation Article 22, Recital 71."

Felzmann, H., Villaronga, E.F., Lutz, C., & Tamo-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, pp. 1-4. [DOI: 10.177/2053951719860542]

Floridi, L., et al. (2018). AI4people—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*, 689–707.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, *38*(3), 50-57.

Hind, M., Wei, D., Campbell, M., Codella, N. C., Dhurandhar, A., Mojsilović, A., ... & Varshney, K. R. (2019, January). TED: Teaching AI to explain its decisions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 123-129).

Hoffman, R.R., Klein, G., & Mueller, S.T. (2020). "The Stakeholder Playbook: An Interim Progress Report." Technical Report from task Area 2, DARPA Explainable AI Program.

Hind, M., Wei, D., Campbell, M., Codella, N. C., Dhurandhar, A., Mojsilović, A., ... & Varshney, K. R. (2019, January). TED: Teaching AI to explain its decisions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 123-129).

IBM (2021). IBM Research Trusted AI [http://aix360.mybluemix.net/consumer]

Johs, A. J., Agosto, D. E., & Weber, R. O. (2020). Qualitative Investigation in Explainable Artificial Intelligence: A Bit More Insight from Social Science. *arXiv preprint arXiv:2011.07130*.

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020, April). Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Klein, G., Hoffman, R.R. & Mueller, S.T. (2022). Naturalistic model of explanatory reasoning: How people explain things to others. In Press, *Journal of Cognitive Engineering and Decision Making*.

Langer, M. (and seven others). (2021). What do we want from explainable artificial intelligence (XAI)?: A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary research. Preprint submitted to *Artificial Intelligence*.

Liao, Q.V., Gruen, D., & Miller, S. (2020). Questioning the AI informing design practices for explainable AI user experiences. *Proceedings of CHI 2020*. New York: Association for Computing Machinery.

Liu, R., and Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics, 42,* 553–568.

Long, W., and Cox, D.A. (2007, June). Indicators for identifying systems that hinder cognitive performance. Proceedings of the Eighth International NDM Conference. K. Mosier & U. Fischer (Eds.). Pacific Grove, CA.

Miller, T. (2017). Explanation in Artificial Intelligence: Insights from the Social Sciences. *ArXiv:1706.07269 [Cs]*. Retrieved from http://arxiv.org/abs/1706.07269

Mittelstadt, B.D., Russell, C., and Wachter, S. (2019). Explaining explanations in AI. In *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, pp. 279–288, Association for Computing Machinery: New York. [doi: 10.1145/3287560.3287574]

Naiseh, M., Jiang, N., Ma, J., & Ali, R. (2020). Personalizing explainable recommendations: literature and conceptualization. In Á. Rocha, H. Adeli, L. P. Reis, S. Costanzo, I. Orovic, & F. Moreira (Eds.), *Trends and Innovations in Information Systems and Technologies* (Vol. 1160, pp. 518–533). Springer International Publishing. [https://doi.org/10.1007/978-3-030-45691-7_49]

Naiseh, M., Jiang, N., Ma, J., & Ali, R. (2020). Personalizing explainable recommendations: Literature and conceptualization. In *Trends and Innovations in Information Systems and Technologies: WorldCIST 2020 Proceedings*, pp. 51 -533. [http://eprints.bournemouth.ac.uk/34805/]

Preece, A., Harborne, D., Braines, D., Tomsett, R., & Chakraborty, S. (2018). Stakeholders in explainable AI. [arXiv: 1810.00184v1].

Ribera, M., & Lapedriza, A. (2019). Can we do better explanations? A proposal of user-centered AI. *Joint Proceedings of the ACM IUI 2019 Workshop*. New York: Association for Computing Machinery.

Russell, S., Jalaian, B., & Moskowitz, A.S. (####). Re-orienting towards the science of the artificial: Engineering AI systems.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems, 10* (4), Article 26. [https://doi.org/10.1145/3419764]

Tate, D.M., Grier, R.A., Martin, C.A., Moses, F.L, & Sparrow, D.A. (2016). "A Framework For Evidence-Based Licensure Of Adaptive Autonomous Systems." Paper P-5325, Institute for Defense Analysis, Alexandria, VA.

Tjoa, E., & Guan, C. (2020). A survey on explainable Artificial Intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*. [https://www.researchgate.net/publication/346017792_A_Survey_on_Explainable_Artificial_Intelligence_XAI_Toward_Medical_XAI]

Tomsett, R., Braines, D., Harborne, D., Preece, A., & Chakraborty, S. (2018). Interpretable to whom? A role-based model for analyzing interpretable machine learning systems. In *Proceedings of the 2018 ICML workshop on Human Interpretability in Machine Learning (WHI 2018).* Stockholm, Sweden.

Wachter, S., Mittelstadt, B., & Floridi, L. (2016). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law., 72, 76-99.

White House. (2019). Executive Order No. 13,859, 84 Fed. Reg. 3967 (February 14, 2019), Sec. 1 [https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/]