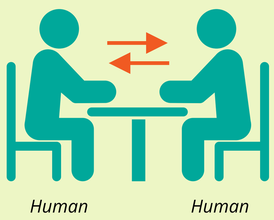
What is an “Explanation”?

What are your thoughts about explanation? How can it encourage clarity of understanding, intent, and information usefulness? Enhance communication?

We learn different ways to *explain* throughout our life. An explanation between two people is a 2-way communication interaction. Explanation is often iterative – a progression toward mutual understanding. It can employ language, non-verbal cues, and relies to some degree on shared contexts (personal and cultural).

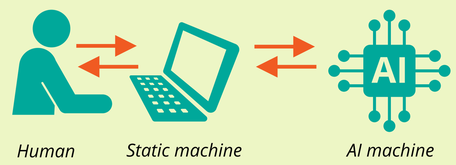


The activities can include building conversational rapport, articulating needs/concerns, context-sharing, validating statements (and broader knowledge) in what is communicated, and seeking clarification. When explanation is successful, it fosters increased understanding and trust.

What can we learn from effective human explanation that can apply to Human-Machine communication? How will it apply to:

Static explanatory messaging (traditional applications/web)

Dynamic AI-supported “explanatory AI” (XAI)



In both cases, systems need to reflect user needs and varying contexts. However, they do so in different ways, and may require different methods and design approaches to formulate the explanations.

What attributes of User Research, Information Architecture and Design can help us create and manage *responsible* explanatory systems?

Types of Explanations

Static ↑ Dynamic

Instructions

Explanations of application/website actions that are based on known functions. These explanations help users understand the operation and sequential steps required by the application.

Actions, challenges

H/M: When a user performs an action, instructional explanation builds awareness of process and requirements, which should aid memory

H: In requirements gathering, people explain work processes and policies, which are then encoded as rules within a system

M: The machine can monitor user actions, including sequence and data accuracy, identifying patterns needing explanation

Error/Warning Messages

Explanations of an anomaly based on an application's internal rules. Error messages describe system or user issue. They ideally explain why (help users understand rules) and give next steps.

Actions, challenges

H/M: Match explanation level to mutual prior interactions/experience

M: Highlight both the location and the nature of errors, focusing user attention and prompting for action

H/M: Increase 2-way communication when it may be unclear whether there is actually an error, rather than an incorrect rule or model

M: Increase explanation depth, quality, if errors are identified as repeated, or user is “stuck”

H: Explanation to elaborate on mental model (particularly for assumed rules)

M: Elicit more user context about task/end goal, to provide different levels and types of information as explanation

Multilingual

Explanations that are available in the user’s preferred language, and translations are accurate although some terms might not translate exactly.

Actions, challenges

M: Since languages have different lengths and typography, depending on the length of explanations there may be space considerations that affect comprehension or usability

M: Identify the languages that are available, and the source language of the information that is used to support the explanation; this can help set expectations when the explanations seem confusing

H/M: In some cases, words or concepts may not be available in the language of choice; users may need to (or decide to) seek information in another language (e.g. “Use the English because there isn’t an Italian equivalent”)

M: Recognize homographs, where the same word may have two meanings, but those multiple meanings are only in one culture, not the language that is in use at the moment

M: Accents may matter, and could reduce user comprehension speed in spoken situations

H/M: Some terms may convey a different sense of urgency, severity or importance in a particular language, which could impact understanding and require different communication moderation, depending on the languages available

Substantive

Evidence, "provability." Focusing on subject-specific information, describing reliability. Possibly referring to AI scoring (the output's fit with different "feature" categorizations).

Actions, challenges

H/M: With traditional search, individual results are given as evidence (with human review = effort), but the ranking algorithm is not transparent; then, human user selections are not “explainable” to the machine, for it to refine its acquisition

M: Visualizing the information space to clarify dominant subject areas and distribution (tightly focused, dispersed, polarized)

M: Subject/sources authority, theories, methodologies, available materials/research

M: Layers of information, such as definitions, “what is…” guidance, typical questions

M: Information scoring, pointing out contrasts/volatility

H/M: How best to explain uncertainty or likely bias?

H/M: Is it useful for AI to ask for next steps? Expect user to request next steps in evidence and explanation?

H: Elaboration or change expressed in their information scope/need

Contextual

Users offer information about contexts that affect their needs. Machine explanation reflects back understanding of user goals, tasks, experience – and clarifies limitations, info depth, etc.

Actions, challenges

M: Elicit/request relevant contexts for an interaction

H/M: Share dimensions that influence decisions; such as people involved, additional medical/health conditions, diagnostics from sensors/other machines, and other H-M interactions

H/M: How might additional context(s) affect rules/algorithms?

M: Confirm (explain) understanding of contexts and which ones matter to the task being done by human and machine; also confirm understanding and impact of any context changes

M: Offer the appropriate information – and only that information – to fit the context

H: How to learn what effect various context aspects impact algorithms and information acquisition?

M: How can the machine effectively/efficiently elicit contexts?

H/M: In medical situations, comparing and if needed combining diagnostic models

Multi-Cultural

The degree to which machines can reflect cultural awareness in explanations. Framing explanations to be user-centered (align with intent) and societally-centered (enhance trust).

Actions, challenges

M: How to identify cues of particular cultural alignment?

M: How general or specific should explanations be? Should they be culturally influenced? Is it possible to not influence culturally? How to be transparent and sensitive?

H/M: What evidence in context expressions reflect a user’s cultural expectations? How much iterative user profiling is needed to assess this?

H/M: “Reading the Air” – a Japanese phrase for sensing and understanding the cultural expectations of other parties in an interaction; being attuned and sensitive to them

M: Setting the right tone: Authoritative, with humility. Help users balance confidence in the machine, confidence in their judgment

H/M: Levels of detail that are welcome, or burdening, or culturally interpreted in unintended ways

Equitable / Intersectional

Explanations avoid inherent biases that can lead to unfair treatment of individuals. Different contextual elements, when combined, could create unexpected concerns among users.

Actions, challenges

M: Unmoderated tone, such as using declarative language in chat, gives a false sense of authority or certainty from an AI system, which risks disempowering people

H/M: Some language “tones” will affect people differently, which risks self-questioning by the user, or even possibly by the AI

M: Consider what communication aspects influence the power dynamics or receptivity of the user, such as timing of delivering an explanation, pace at which it is delivered (fast or slow), interruption or over-talking as part of turn-taking in conversation

M: Communicating in ways that might be perceived as dismissive of the recipient (such as a style perceived as “mansplaining”)

H: Having the feeling of “missing out” because information is excluded or described as not relevant to the user (such as patients not being given access to content because it is meant for doctors/nurses)

M: Language, vocabulary, tone can affect sense of “us” or “other”

Relational

Human-AI interaction now, and in future, will be longitudinal. Knowledge of each other, previous interactions, and changing experience/expectations need to be expressed and explained.

Actions, challenges

H/M: Longitudinal (multiple sessions over time) rely on familiarity (and a sense of memory/history) between the user and system. An expectation of updating understandings should be in any model

H/M: How should changes in learning, experience, intents, contexts, models, and the information space be shared?

H: In diagnostic decision support, the ongoing condition and treatments will be evolving. This requires very current, shared contextual information

H/M: In conversations, are multiple parties involved over time?

M: How “familiar” should a machine be? Should there be a move from low-context to higher-context communication styles? Should language use/form change with knowledge?

H: Accurate, consistent explanations engenders trust. What else is a key to supporting trust?

Evolving

Throughout the life of an AI system, it evolves as data changes, interactions are refined, models vary, and human expectations change. Evolution must be continually monitored/explained.

Actions, challenges

M: Routine diagnostics and proactive AI explanations can signal evolution in the info space, internal models, or human uses

H: Developers, data scientists and UX must validate and explain any type of change, however small, in a system

M: Certain types of data benefit from visualization: trend data, scoring information features, statistics of human use/ responses, problems with human request input, etc.

H/M: Iteration can lead to evolution – how does an AI system explain internal ecosystem iterations or model drift?

H: Can competing models (between different internal agents) be identified? How would agents provide explanations/evidence?

M: How are potential paradigm shifts explained… When emerging evidence begins to diverge from “known facts”?

Direct, Rule-Based ← → Complex, Interactive

Design for Context
Making software applications and web sites easy to use

Prepared by Duane Degler for the IA Conference ’23.
Shared 3.30.2023.

OXIDE.AI