RED FLAGS CATEGORY

Product Design, Development, and Maintenance

Ford Foundation

11

RED FLAG

The product claims to be completely new, “disruptive,” or different in all relevant facets.

AT A GLANCE

- Products claiming to be completely new or "disruptive" should be scrutinized, as they may not consider past lessons learned in the field.
- The vendor's solution may create new harm in other domains
- To identify this red flag ask about direct and indirect stakeholders, who are the vendor’s competitors, and how is the proposed product or technology similar and different from existing solutions.

In most cases, a product that claims to be completely new or different from all precedents is not completely new, especially in the public sector. If a vendor claims to provide a groundbreaking solution to a public interest issue, it is possible the vendor has not engaged with the existing domain and current players.

The vendor may not build upon lessons that past products/vendors have learned. Thus, the vendor may repeat or amplify past mistakes. Moreover, the vendor’s solution may be solving an issue in one domain, but doing so by creating new harm in another.

EXAMPLE

A potential vendor may claim to predict how well a child will perform in school by feeding a large amount of personal data into an AI model. Prior research has shown that in reality, even large machine learning models with access to fine-grained data collected over years for each child is unable to outperform a simple regression model using a few data points.[1] In this case, the use of AI can be an excuse to access/collect more data.

QUESTIONS TO IDENTIFY THIS RED FLAG

Who are the current stakeholders and parties involved in this domain?

Who are the key competitors and how does this product differ from them?

How is the proposed product or technology similar and different from existing solutions?

RESOURCES

- How to recognize “AI Snake Oil”
- On NYT Magazine on AI: Resist the Urge to be Impressed

[1] Salganik, Matthew J., Ian Lundberg, Alexander T. Kindel, Caitlin E. Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M. Altschul et al. "Measuring the predictability of life outcomes with a scientific mass collaboration." Proceedings of the National Academy of Sciences 117, no. 15 (2020): 8398-8403. https://www.pnas.org/doi/full/10.1073/pnas.1915006117

12

RED FLAG

The product replaces an existing product with an interface that is very different from the one that users are accustomed to or the user interface/design is inaccessible to people with disabilities or intimidating to those lacking technical or digital literacy.

AT A GLANCE

- Significant changes to a user interface can cause confusion and limit accessibility, especially for those with limited technical or digital literacy.
- To identify this red flag ask whether the product has been tested with a range of users with varying degrees of digital fluency and disabilities, what are the potential consequences of accidental misuse, and how does the design differ from what users are accustomed to.

Significant changes to an interface, especially one that users have grown used to, can cause confusion or open the door to inadvertent misuse. Since large changes in UI often assume specific technological literacies, these changes can make technology more difficult to use for clients, limiting accessibility.

EXAMPLE

A prison system migrates from a PC-based communication system to a tablet-based one. Further, incarcerated people use the communication system to call family members. It is possible that, for instance, those currently incarcerated are allowed to use the device for a limited period of time. If they are not used to the tablet interface and controls, it is possible that it will take them much longer to navigate the system, reducing the amount of time for the call itself. Even logging into the system may be a challenge if there are people who remember passwords via muscle memory on a traditional keyboard.

QUESTIONS TO IDENTIFY THIS RED FLAG

How does the design differ from the interface that clients are accustomed to using?

What user testing has been done to show that the design is accessible; namely have you tested the product with those with various degrees of digital fluency?

What are ways that the technology could be accidentally misused? If the interface is inadvertently misused, what is the range of possible consequences? Is it possible to remediate the consequences?

How do you plan to overcome mistrust or unfamiliarity in order to increase adoption and impact?

RESOURCES

- User Interface Design for Low-literate and Novice Users: Past, Present and Future
- Making the Web Accessible: Strategies, standards, and supporting resources to help you make the Web more accessible to people with disabilities.
- Protecting Older Users Online
- Building Access: Universal Design and the Politics of Disability

13

RED FLAG

The proposed project does not sufficiently follow industry best practices including security, privacy, openness, interpretability, and non-discriminatory design.

AT A GLANCE

- Products should follow technical industry best practices.
- Products should be tested for stability, encryption, and resistance to cyber-attacks for cybersecurity, and adhere to "privacy by design" principles for privacy.
- Tools used by public agencies should be audited for discrimination, interpretability, and accountability, and vendors should be transparent about the results of these audits and their decision-making processes.

From a technical standpoint, there are several industry standards against which developers can test products. Testing is very much dependent on the type of technology. We understand that philanthropic organizations and government agencies may not have the in-house expertise to fully run all the necessary tests. However, having general knowledge about those testing criteria is necessary. In addition, organizations can work with external experts to be able to test product/service performances based on the relevant standards.

For cybersecurity, products should be tested for stability, authentication, encryption, and resistance to cyber-attacks. For privacy, products should adhere to "privacy by design" principles including minimal data collection, privacy-by-default settings, and retaining data as long as needed. For human-rights-centric UX/UI[1] design, in addition to security and privacy, products should be tested based on accessibility criteria such as network and device quality, beneficiaries' digital literacy levels, physical and mental impairment, etc.

Furthermore, computational tools that help public agencies to make decisions about certain applications (e.g., predictive risk assessment tools in child welfare practices, student assignment algorithms for public schools) often rely on historical and demographic information.

Researchers have shown that these tools are prone to discrimination based on gender, race, religion, and other socioeconomic factors.[2] Apart from assessing the more complex and long-term impacts of these tools, these tools should also be audited based on other technical criteria. For instance, issues around over and under-representation arise during the process of collecting and annotating data that is used to train, validate, test, and optimize the system.

In addition, these systems are prone to making "unfair" decisions based on the input variables (whether directly about protected groups such as race, religion, or age, or proxies such as zip code, phone area code, education level) and statistical models that are selected during the design and development process. Vendors may also use more complex technical methods to design these systems.

When the system becomes very complex, there might be no transparency in how that system makes a certain decision (e.g., why this tool thinks that family A should be denied access to welfare benefits but not family B; why asylum seeker A's application should be granted but not B's).

A lack of interpretability in these systems may lead to confusion and weaken beneficiaries' ability to hold public agencies to account. When mistakes are made, there is no clear answer to who should be blamed: the tool, the vendor, the public employee, or the agency?

QUESTIONS TO IDENTIFY THIS RED FLAG

What industry best practice standards did you use to design and test your product? Use the following resources and ask about privacy, security, fairness, interpretability, accessibility, openness, and sustainability.

Can we or our trusted technical partners test your product? Do we need to sign an NDA for it? If yes, why and what does it include?

Is your product documentation available publicly? If not, what prevents you from keeping the documentation open?

Do you perform tests to determine whether your product or tool is creating discriminatory, adverse outcomes for certain demographic groups? If so, how do you obtain the demographic data in order to perform these tests?

Can you share your audit and/or impact assessment reports? Who conducted the audits? Ask whether the audits have been conducted by the company itself, by consultants who were commissioned by the company, or externally by advocates, technologists, and researchers.

RESOURCES

For Privacy and Security

- The Digital Standard by Consumer Reports
- The Open Web Application Security Project or OWASP's Mobile Security Testing Guide
- The OWASP's Testing for Weak Encryption
- The Mozilla Observatory
- Privacy by Design: The 7 Foundational Principles
- Security Planner

For UX/UI Design

- Digital Security and Privacy Protection UX Checklist

For Algorithmic Fairness

- White House Blueprint for an AI Bill of Rights
- Microsoft Fairlearn toolkit
- IBM Fairness 360
- Google's "What If?" Tool
- Eticas Guide to Algorithmic Auditing
- Other tools

RESOURCES

For Algorithmic Fairness (Continued)

- Datasheets for Datasets
- Model Cards for Model Reportinga

For Algorithms Interpretability and Explainability

- Introduction to Interpretable Machine Learning (I, II)
- AI Explainability 360, IBM

For Sustainability

- Principles of Green Software Engineering

For Openness

- Critical Digital Infrastructure

Other Resources for Responsible Product Design

- Value Sensitive Design: Envisioning Cards
- AI in Education Toolkit for Racial Equity: How to mitigate racial bias in the design and development of your product
- AI risk management framework
- Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance

[1] User Experience/User Interface
[2] Eubanks, Virginia. Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press, 2018.

14

RED FLAG

Vendor is not able to explain how the product/service works in an accessible manner, without using technical terms.

AT A GLANCE

- Excessive use of technical terms and acronyms without clear explanations can indicate that a vendor is over-selling their product.
- Inability of the vendor to explain the product can result in public agencies also being unable to understand it, leading to accountability issues.

Vendors should be able to explain how their tool works in an accessible manner. During our interview with digital rights advocates and social entrepreneurs, it became clear that throwing in terms and acronyms such as AI, algorithm, machine learning, deep learning, blockchain, etc. without being able to simply explain why that technology is used in a system is an indicator of a vendor over-selling its service.

In addition, if a vendor is not able to explain its service, public agencies that use the service will not be able to either. This will lead to a further lack of accountability.

EXAMPLE

A civic tech company proposes developing a method for faster and safer political participation such as voting. Their method relies on blockchain technology. During their conversation with funders, they use terms such as “blockchain,” “public ledger,” “private ledger,” “chain,” etc. without elaborating what they mean by the terms and why this technology is relevant.

QUESTIONS TO IDENTIFY THIS RED FLAG

Can you explain how your tool works for users and stakeholders with a lower level of technical literacy? If not, what is limiting you from doing so?

15

RED FLAG

The product locks you in and/or is not easily repairable.

AT A GLANCE

- Products that lock in users and are not easily repairable can cause problems in switching to another better product.
- To identify this red flag ask about compatibility of the product with standard data formats, proprietorship of the product, issues with rolling out updates, repairability of services, and processes for minimal disruption to end users in case of technical issues.

Sometimes products are designed in such a way as to make switching to another (better) product costly and resource-intensive. Sometimes a tool may require proprietary data formats that are only compatible with that certain software. This becomes a serious issue when a vendor and public agency work together to pilot a service. After the pilot phase, it may be too costly for the agency to switch to other services. They may simply decide to continue working with that vendor because they are "locked in" with them. Proprietary software can also make it difficult for public agencies to repair a system or service.

During the course of our interviews, several entrepreneurs and advocates mentioned that these issues can arise due to the lack of interoperability among systems. There is no doubt that the interoperability of digital systems is important; after all, the Internet is built on the principle of interoperability and seamless information exchange.

However, from an anti-surveillance perspective, interoperability of data sharing systems between and within governments, without adequate safeguards, may result in harmful consequences. An example could be frictionless data sharing practices between police departments, Immigration and Customs Enforcement agencies, and other public offices that are involved in managing education, health care, and welfare services

In this example, this may result in surveilling refugees and immigrants, arbitrary arrest, and denying them access to public spaces/services.

EXAMPLE

A technology vendor wins a bid to develop custom-built data management systems for a county. A few years ago, a city in that county updated its data management system. However, the city's system is not compatible with the one for the county. To solve the issue, the vendor proposes to update the city's internal system as well.

In addition, the vendor proposes to sell other custom-built add-on services (project management system, internal messaging platform, invoice management, etc.). This pattern repeats itself every time the county, city, or state needs to upgrade its digital infrastructure. The chaotic situation hinders public officers' services while exposing sensitive government data to instability and malicious activities.

QUESTIONS TO IDENTIFY THIS RED FLAG

If your product needs data as its input, what kinds of standard data formats is it compatible with? Can the data be exported to similar products?

What proprietorship do you have on this product?

Have you experienced any issues with rolling out updates to your users – either end users or public agencies?

How can public agencies repair your services? Do they need to keep in touch with you as long as they use this service?

What processes do you have in place to ensure minimal disruption to end users/affected communities in the event of technical issues?

RESOURCES

Digital rights to repair act

16

RED FLAG

Developers don't make explicit how a certain technology or product will be maintained or adapted in the future.

AT A GLANCE

- A lack of explicit maintenance plans for technology or products can lead to a lack of sustainability, making it harmful in the future if the environment, ecosystem, or regulations change without updates.
- To identify this red flag ask about the frequency of updates for the technology and training data, preparation for seamless transfer of services, and willingness to be held accountable for maintaining a certain quality of service.

Without making explicit how a certain technology or product will be maintained or adapted in the future, vendors can inadvertently create a product that is helpful today but harmful in the future.

Without stating how a technology will be maintained in the future, vendors indicate a lack of sustainability. Technology can become faulty and harmful if the environment, ecosystem, or regulatory landscape changes without updates to the technology.

EXAMPLE

A research team at Vanderbilt University was able to show that a model trained to predict hospital mortality rates using data from 2006 deteriorated in quality over time. One model in particular provided a mortality rate prediction for 2013 when the actual observed value differed by 25 percent from the predicted value. Because of shifts in the cases that hospitals in the areas were treating the original model had gone stale.

QUESTIONS TO IDENTIFY THIS RED FLAG

How and how often will the technology and/or training data be updated once it is released?

Can you describe your client training program? How do you help your clients have direct access to you or become independent in maintaining and troubleshooting the service?

Are you prepared to work with a future vendor to seamlessly transfer services without interruption? How?

Would you be open to contract's violations terms that impose fines if you don't maintain a certain quality of service threshold?

RESOURCES

- A primer on AI model drift
- The Maintainers