

UPDATE

Ethical Web

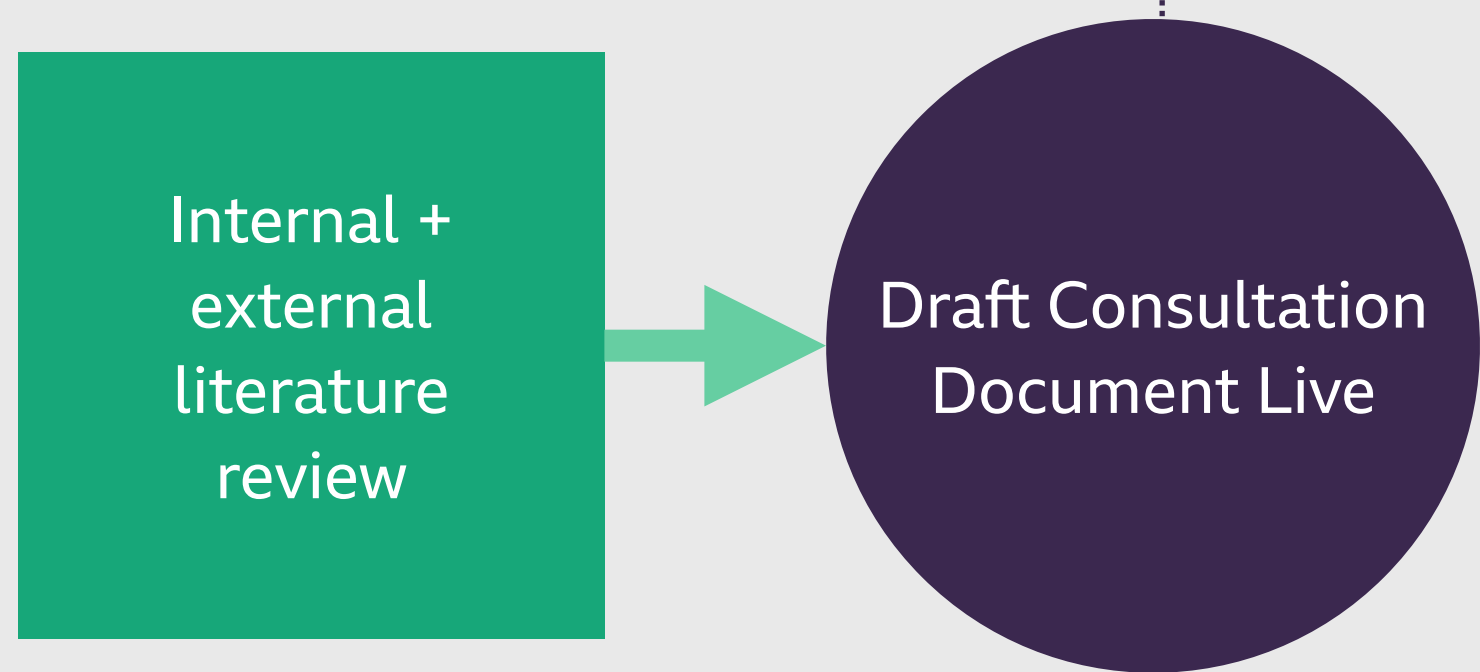
Machine Learning

James Fletcher - BBC Lead, Responsible AI/ML

10 March 2021

W3C Web ML: Ethical considerations

Now



Draft Consultation Document

<https://docs.google.com/document/d/1n55liw3cAcrlDMLvRPEAdV1ANWT9QzgOZ6R0pUaSVY4/edit?usp=sharing>

DRAFT CONSULTATION DOCUMENT: W3C | Ethical Web ML Working Group Note

Status of this document

This is a document to enable consultation and stakeholder input into the Web Machine Learning Working Group's Note on Ethical Principles for Web ML.

As such, it contains material documenting and supporting certain decisions which is relevant for effective consultation, but might not ultimately end up in the final working group note.

It also contains background information aimed to help those unfamiliar with ethics orient themselves, both generally and with regard to ML ethics specifically.

Views are welcomed on how much of the rationale and background should end up in the final doc, and how that is best achieved (e.g. in Appendixes)

It is also a draft / work in progress provided to enable input as early as possible. Some sections are complete drafts, while others are still incomplete - the latter are marked WIP.

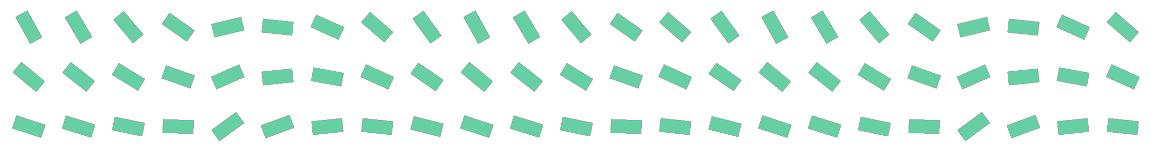
Abstract

This document discusses ethical issues associated with using Machine Learning and outlines considerations for web technologies that enable related use cases. It proposes a set of ethical principles to guide use of Web ML, and an approach to effectively turn principles into practical consideration of harms, risks and mitigations for specific use cases.



Arriving at the principles

- Universality
- Alignment with W3C Values & Principles
- Recommending use of UNESCO Values & Principles



UNESCO Values & Principles

The UNESCO Values are:

- Respect, protection and promotion of human rights and fundamental freedoms and human dignity
- Environment and ecosystem flourishing
- Ensuring diversity and inclusiveness
- Living in peaceful, just and interconnected societies

The UNESCO Principles are:

- Proportionality and Do No Harm
- Safety and security
- Fairness and non-discrimination
- Sustainability
- Right to Privacy, and Data Protection
- Human oversight and determination
- Transparency and explainability
- Responsibility and accountability
- Awareness and literacy
- Multi-stakeholder and adaptive governance and collaboration

W3C Web ML: Ethical considerations

Now



Other areas for comment & input - Guidance

Principles		
Proportionality and Do No Harm	<p>The choice to use AI systems and which AI method to use should be justified in the following ways: (a) the AI method chosen should be appropriate and proportional to achieve a given legitimate aim; (b) the AI method chosen should not infringe upon the foundational values captured in this document, in particular, its use must not violate or abuse human rights; and (c) the AI method should be appropriate to the context and should be based on rigorous scientific foundations. In scenarios where decisions are understood to have an impact that is irreversible or difficult to reverse or may involve life and death decisions, final human determination should apply. In particular, AI systems should not be used for social scoring or mass surveillance purposes.</p>	<p>TAG EWP: When we are adding a feature or technology to the web, we will consider what harm it could do to society or groups, especially to vulnerable people. We will prioritize potential benefits for web users over potential benefits to web developers, content providers, user agents, advertisers or others in the ecosystem, in line with the priority of constituencies. We will ensure the requirements and views of marginalized communities and underrepresented groups are heard and respected.</p>
Safety and security	<p>Unwanted harms (safety risks), as well as vulnerabilities to attack (security risks) should be avoided and should be addressed, prevented and eliminated throughout the life cycle of AI systems to ensure human, environmental and ecosystem safety and security. Safe and secure AI will be enabled by the development of sustainable, privacy-protective data access frameworks that foster better training and validation of AI models utilizing quality data.</p>	<p>TAG EWP: When we add features to the web platform, we are making decisions that may change the ability of people to protect their personal data. This data includes their conversations, their financial transactions and how they live their lives. We will start by creating web technologies that create as few risks as possible, and will make sure people understand what risks they are taking when they use the web.</p> <p>Security is subject to its own process and horizontal review, see: Self-Review Questionnaire: Security and Privacy</p>
Fairness and non-discrimination	<p>AI actors should promote social justice and safeguard fairness and non-discrimination of any kind in compliance with international law. This implies an inclusive approach to ensuring that the benefits of AI technologies are available and accessible to all.</p> <p>AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems. Effective remedy should be available against discrimination and biased algorithmic determination.</p> <p>Furthermore, digital and knowledge</p>	<p>TAG EWP: Anyone should be able to meaningfully participate in the creation of specifications, user agents, and content, and the platform should enable a fully accessible end user experience. We will build internationalization and localization capabilities into our specifications and websites. We will accommodate people on low bandwidth networks and with low specification equipment. The web platform and the tools we use to create it must be accessible to people with disabilities, including visual, auditory, physical, speech, cognitive, language, learning, and neurological disabilities.</p> <p>Accessibility is the domain of the Accessible Platform Architectures (APA) Working Group, and subject to its own process and horizontal review: see draft</p>

Other areas for comment & input - Issues & Case Studies

3. (WIP) General ethical issues in Machine Learning

(NOTES:

- This section is incomplete draft - content for illustrative purposes, but needs finishing and inclusion of examples
- For discussion how much of this background we need)

For a general background on ethics and its relevance to ML, see [Appendix 1. Background: Ethics & Machine Learning](#)

So what sorts of ethical questions does machine learning raise? Based on case studies of where ML has gone wrong, there are a number of key issues.

Accuracy

One concern is how accurate the predictions of ML systems are. This is increasingly of concern where ML is used to determine eligibility for things we might consider essential, such as healthcare or welfare payments, or where it is determining outcomes in areas of key human rights such as justice systems. (LINKS TO CASE STUDIES TO BE ADDED) There is also a more general concern about the [over-hyping of the ability of AI to predict certain things at all](#).

Bias & Fairness

Related, and perhaps the most prominent ethical concern about ML, is bias. Bias has a number of meanings, but the key concern here is where the application of ML leads to negative outcomes (such as inaccurate predictions) which disproportionately affect specific **groups**, particularly where those groups align with protected characteristics such as race, gender,

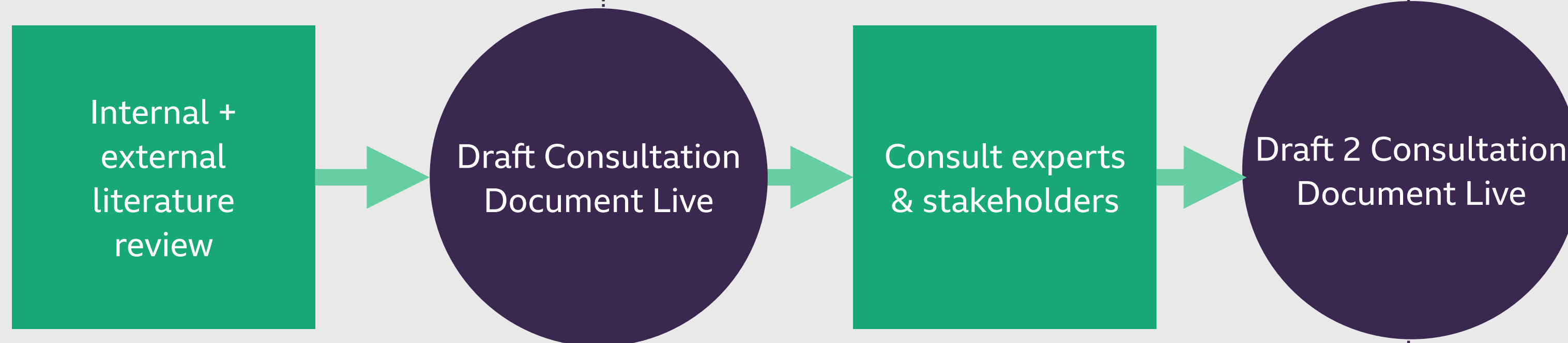
Other areas for comment & input - High Level Risks / Mitigations

Principle	Risks	Possible Mitigations
Fairness and non-discrimination	<p>Scaling up ML via browsers creates risks of scaling up bias issues linked to ML training.</p> <p>ML approaches optimize for the majority, leaving minorities and underrepresented groups at risk of harm or sub-optimal service (see e.g. Treviranus)</p>	<p>Browser-assisted mechanisms to find out about the limitations and performance characteristics of ML models used in a Web app. This would build on an approach published in Model Cards for Model Reporting where making this report machine-discoverable would allow for the web browser to offer a more integrated user experience.</p>
Proportionality and Do No Harm		
Safety and security		
Sustainability		
Right to Privacy, and Data Protection		
Human oversight and determination		

W3C Web ML: Ethical considerations

Now

~21 Mar



- **Guidance**
- **Issues & Case studies**
- **High level risks and mitigations**

W3C Web ML: Ethical considerations

