Attachment B – Design Actions

Determine actions the system can take

AI systems may be designed to:

- Make decisions directly,
- Produce recommendations or predictions to be assessed by human experts, or
- A combination of both.

The system may also be designed to make additional actions like acquiring more data.

All of these potential decisions, recommendations and actions need to be designed, in consultation with experts, to ensure the defined objectives and benefits of the AI system are realised and identified risks are minimised.

Design interfaces

Individuals may interact with the AI-enabled system directly, for example, by entering data about themselves into a web interface (e.g. loan or credit card applications), or the system may support human operators (e.g. customer service centres fielding enquiries and requests).

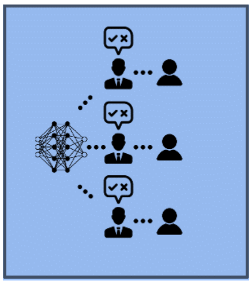
Designing the appropriate interface requires collaboration with the appropriate experts including data scientists designing the inference system, user experience (UX) designers, and representative end-users. Note that some of the design decisions might not be able to be confirmed at this stage and need to be determined during the testing.

The design process also needs to cover user training considerations, interface design for monitoring the AI system, and special considerations for users with disabilities and accessibility needs.

Agree on the role of human decision-making

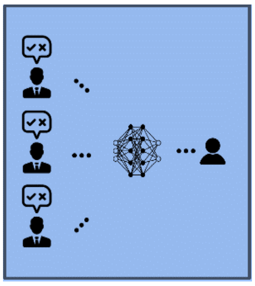
The most role of human decision-making in the AI-enabled system may need to be determined through experimentation and testing. It may not be necessary or even beneficial for a human to oversee every machine decision. Some examples of human/machine decision-making relationships are shown in the following table:

Configuring the right human/machine oversight is highly dependent on the problem being addressed



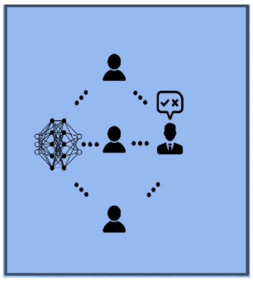
Humans check every decision

(e.g. humans checking AI diagnosis before recommending invasive surgery)



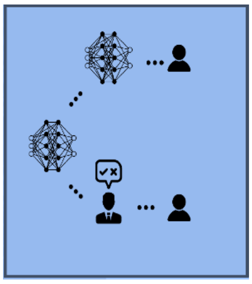
Machine mediates human experts

(e.g. machine find commonalities and deviations in past decisions by different customer service staff)



Humans check contested decisions

(e.g. humans consulting with all experts across various agencies on AI decision to devise the right intervention program)



Machine selects between human and machine decision

(e.g. an image labelling algorithm that sends images it is uncertain of to human experts)

Create operating parameters

Agencies need to define the operating parameters for the AI system to ensure it meets its objective. Examples include:

- “probability thresholds” determining whether to act or not act;
- “uncertainty thresholds” that invoke a different decision process; and
- “weights prioritising” different metrics in the system objective.

For example, a system that selects individuals for a social services intervention can have parameters that control:

- How readily the intervention is given out based on a likelihood of effectiveness
- Relative importance of different considerations for selection
- Threshold of uncertainty beyond which a human caseworker should make the decision
- Degree to which random assignments should be done to gather experimental data

Agencies can subsequently change these parameters to:

- Achieve different balance between competing objectives
- Place constraints on actions
- Control the structure of the oversight and human decision-making interface as previously described.

It is important that a procedure for review and redress is built into the design of the system for potentially erroneous or contested decisions.

It is also important to test the interface on a representative sample of the intended users to capture their interactions with the system and their feedback. This is especially important when the users may have special needs or disabilities.