Human Factors in AI

⚠ AI systems are highly susceptible to

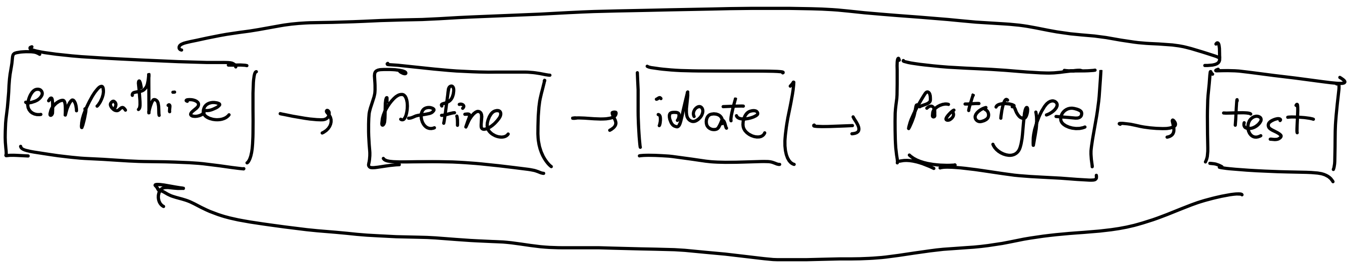
- infringing privacy
- biased decision making
- resistance to adoption

⚠ Design thinking is a human-centered methodology for creative problem solving

3 E's of design thinking

- Empathy
- Expansive thinking
- Gxperimentain

⚠ Design thinking process



⚠ empathy → set aside your own assumptions and gain insight into your user's needs

⚠ define → synthesize the information collected in the empathize stage

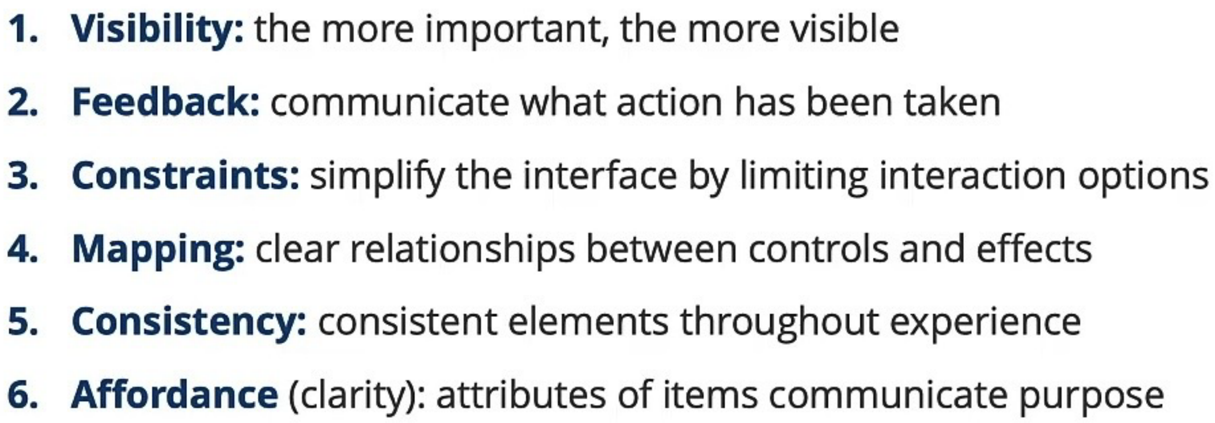
⚠ ideate → generate ideas of ways to solve the problem

⚠ Prototype should be quick and cheap
- ↳ to answer questions / test hypothesises

⚠ Task Analysis
- ↳ analysis of the step-by-step details of a user's task flow
- ↳ why? → ensures understanding of user's problem
  - ↳ reduced bias from preconceived assumptions
  - ↳ provides insights into steps that can be improved
- ↳ How? → observation - trigger, goal, decisions, tools
  - ↳ diagram the task flow

⚠ UX Design Principles

↳



1. **Visibility:** the more important, the more visible
2. **Feedback:** communicate what action has been taken
3. **Constraints:** simplify the interface by limiting interaction options
4. **Mapping:** clear relationships between controls and effects
5. **Consistency:** consistent elements throughout experience
6. **Affordance** (clarity): attributes of items communicate purpose

⚠ User Inputs

- forms , uploads, votes / ratings , actions
- used to collect data from users

⚠ Cold start problem

- if we are relying on user-supplied data for our model, we may initially not have enough to build a quality model

⚠ Transparency considerations in AI

- where AI exists / what it does
- what data it uses
- How it reaches its output
- limitations

⚠ How to provide transparency

- cite data sources / attributes used
- give insight into importance of attributes
- provide basis for model output

⚠ feedback loops

- many ML systems employ feedback loops where user interactions with a model influence the outputs they see over time
- can be explicit & implicit
  - based on direct user feedback
  - based on user actions as a result of a model

⚠ Data privacy

- right of users to have control over how their information is collected, used and shared

⚠ Fair Information Practices (FIPs) organized into 4 themes:

- rights of individuals
- controls on information
- information life cycle
- management of personal identifiable information (PII)

⚠ why protect user data privacy

- avoid violation of privacy laws (GDPR, HIPAA, ...)
- gain trust of users
- maintain reputation

⚠ How to protect user data privacy

- compliant policy and practices
- Privacy by design
- technological approaches

⚠ Technological approaches to protect privacy

- federated learning
    - allows users/devices to contribute towards improving a shared model without sharing their data
- differential privacy
    - calculation/modeling approaches where one can not tell from the output whether any indivisual's data was included in the input dataset

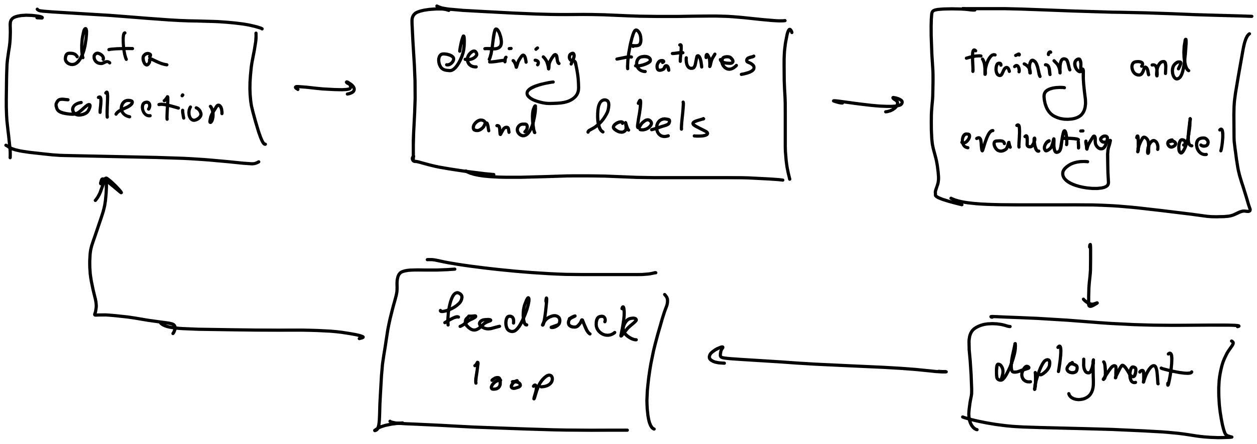
⚠ Ethical risks of AI

- allocate harm
  - opportunities or resources are withheld from certain people / groups
- representational harm
  - certain people/groups are stigmatized or stereotyped

⚠ Three criteria of ethical AI systems

- fair
- accountable
- transparent

⚠ Sources of bias



⚠ Types of bias

- algorithmic
- historical
- representation
- measurement
- learning
- deployment
- feedback loop

⚠ Tools to mitigate ethical risk

- datasheets for datasets
- ethical checklist
- ethical pre-mortems

⚠ objectives of a dataset datasheet

- for dataset creators
  - encourage best practices in collecting data
  - foster reflection on risks and implications of use
- for dataset consumers
  - provide transparency to support decisions on whether / how to use dataset
- for users of models
  - contribute to explainability of model outputs

⚠ Defining fairness goals

- define groups of significance
- determine what "fair" means

⚠ Anticipation of fairness issues is key to mitigation

⚠ Artificial general intelligence (AGI)

└→ ability of an intelligent agent to learn any intellectual task that a human can

⚠ Narrow AI

└→ ability to accomplish specific pre-learned problem solving tasks

⚠ automation → replacing humans

augmentation → supporting humans

⚠ forms of AI augmentation

├→ triage

└→ decision support

⚠ Intentional focus on building model trust and proper onboarding can ensure adoption