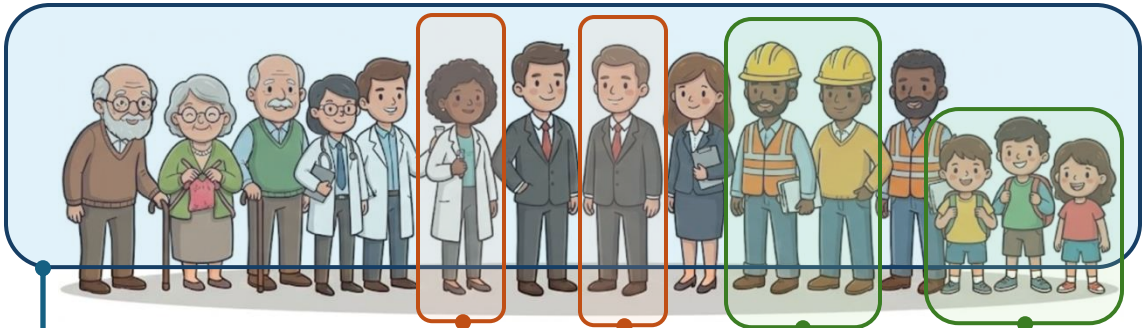

Aggregate Preference

Universal Value
- Fairness
- Honesty
- Safety
- ...

Social Norm
- Quietness
- Decency
- Order
- ...

Individual Preference

Personal Style
- Tone
- Taste
- Pace
- ...

Content Preference
- Density
- Logic
- Utility
- ...

Community Preference

Shared Interest
- Technology
- Gaming
- Arts
- ...

Collective Objective
- Consensus
- Stability
- Welfare
- ...

Preference Types

One-size-fits-all

Global Model

Individual-level

User Model A
User Model B
User Model C
User Model D

Community-level

Community Model A *(Gender)*
Community Model B *(Occupation)*

Alignment Approaches