Trustworthy Digital Society

Joint research initiative of UNSW & UTS

Trustworthy Digital Society

Youth, Trust & Technology

In what ways can young people's participation in the governance of AI-generated images and deepfakes foster civic, technological, and cultural trust in digital society?

The project explores youth-led pathways for governing deepfakes through participatory research that connects governance, technology, and culture as interdependent spaces of digital trust.

Lead Chief Investigators:

Keith Heggart
Catarina Pinto Moreira

Chief Investigator:
Lena Petrova

BACKGROUND

The spread of AI-generated images and deepfakes has eroded trust in institutions, technology, and civic culture. While technical safeguards
such as watermarking are emerging, these alone cannot rebuild public confidence. The YTT project recognises that trust must be restored collectively—through civic participation, education, and youth-led innovation.

SIGNIFICANCE / WHY IT MATTERS

Deepfake harms are not only technological but cultural and systemic. Current solutions often reflect neoliberal logics of self-responsibility, overlooking shared accountability. The YTT project reimagines trust as a collective practice—one that extends beyond detection tools toward civic participation, education, and moral responsibility.

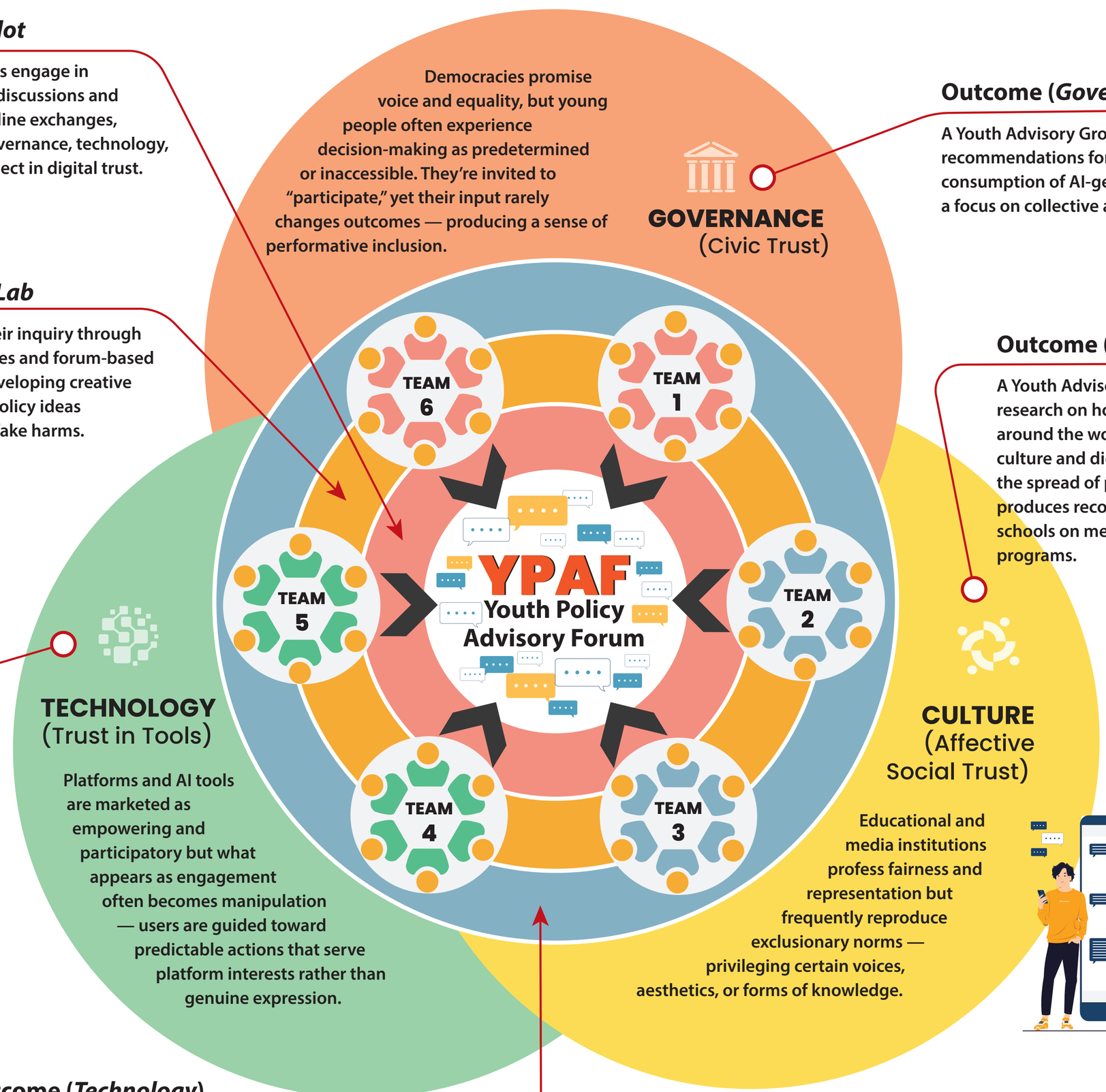
The YPAR will first be trialled with UTS and UNSW students in 2026, before being implemented in schools in 2027.

Phase 1 — *Pilot*

Youth participants engage in fortnightly team discussions and asynchronous online exchanges, mapping how governance, technology, and culture intersect in digital trust.

Phase 2– *CoLab*

Teams extend their inquiry through co-design activities and forum-based collaboration, developing creative prototypes and policy ideas addressing deepfake harms.



Outcome (*Governance*)

A Youth Advisory Group develops a set of recommendations for the safe consumption of AI-generated images, with a focus on collective accountability.

Outcome (*Culture*)

A Youth Advisory Group conducts research on how young people around the world use popular culture and digital spaces to counter the spread of post-truth logic, and produces recommendations for schools on media competency programs.

Outcome (*Technology*)

A Youth Advisory Group researches the effectiveness of watermarking, labelling, and other fake-detecting technologies.

Phase 3 — Policy & Media Literacy Framework

Findings from earlier phases are consolidated into actionable recommendations for schools, educators, and policymakers.

Transforming Distrust into Civic Agency