

Use Case

- lots of sites where image/video contents can't be readily translated
- putting an "AR" layer over the digital environment, track the screen contents, try to identify contents and their contained texts
- translate texts *immediately* (automatically or on button click)

USP

- Layer System: Match text position of translations with original text positions
- If on Apple/Google Store: direct competitors are e.g. Screen translator
 - Workflow there: Open -> Close -> Scroll -> not the best UX
 - general principle is the same, but UX sucks

Key Questions

- focus on these questions planned for the first **2-3 weeks**

1) How can we detect world elements properly?

- 1) Using an API?
- 2) Developing your own AI?
- 3) Any (other) idea?

2) What kind of text detection APIs are out there on the market?

- 1) Alternatives to Google's Vision AI
- 2) Analysis list on this topic: advantages/disadvantages, open source or commercial

3) Problem with Google's Vision AI: Google has problems distinguishing contents by context

- 1) How could we build our own Text Recognition AI?
- 2) What do we have to consider here theoretically?
- 3) What kinds of pre-defined models exist?
- 4) What kind of data/things do we need to train better?
- 5) How can contextual partitioning be achieved?

4) What technology can we use to stabilize the layer on an identified element? (text is tracked, above translation is moved according to scroll input fluently) Possible Identifying and Stabilizing solutions

5) How to distinguish a good text detection system from a bad one?