

Easy, Reproducible and Quality-Controlled Data Collection with CROWDAQ

Qiang Ning¹ Hao Wu² Pradeep Dasigi³ Dheeru Dua⁴ Matt Gardner³
Robert L. Logan IV⁴ Ana Marasović³ Zhen Nie²

¹Amazon (work finished at the Allen Institute for AI) ²Hooray Data Co., Ltd. ³Allen Institute for AI ⁴University of California Irvine

Abstract

High-quality and large-scale data are key for many AI systems.

Three common challenges

- Designing a user-friendly annotation interface
- Training annotators efficiently
- Dataset reproducibility

CROWDAQ addresses issues above

- Customizable UI construction (no frontend expertise required)
- Annotator qualification management via user-defined multiple-choice exams
- Standardized data collection pipeline that one can use to simply “pay and collect”

CROWDAQ is **open-source** and provisioned by **Hooray Data** for free.

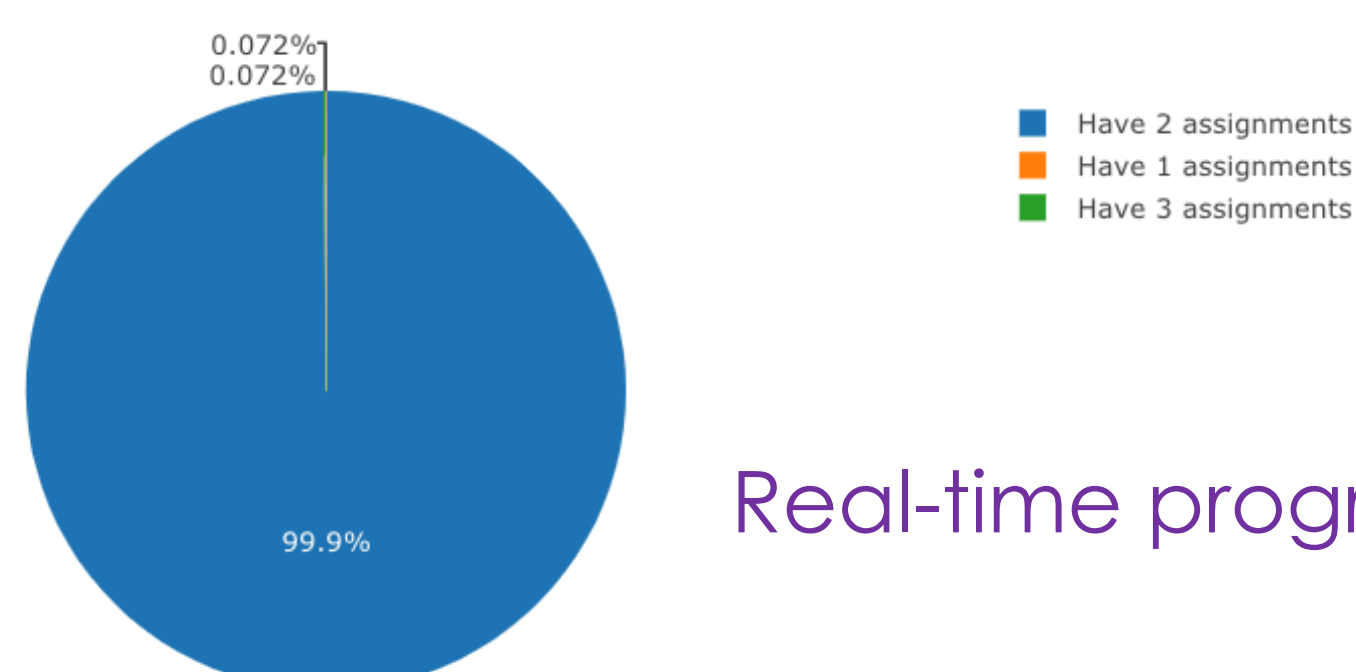
Access CROWDAQ

<https://www.crowdaq.com/>

Demo user: emnlp-demo
Password: **crowdaqisawesome**

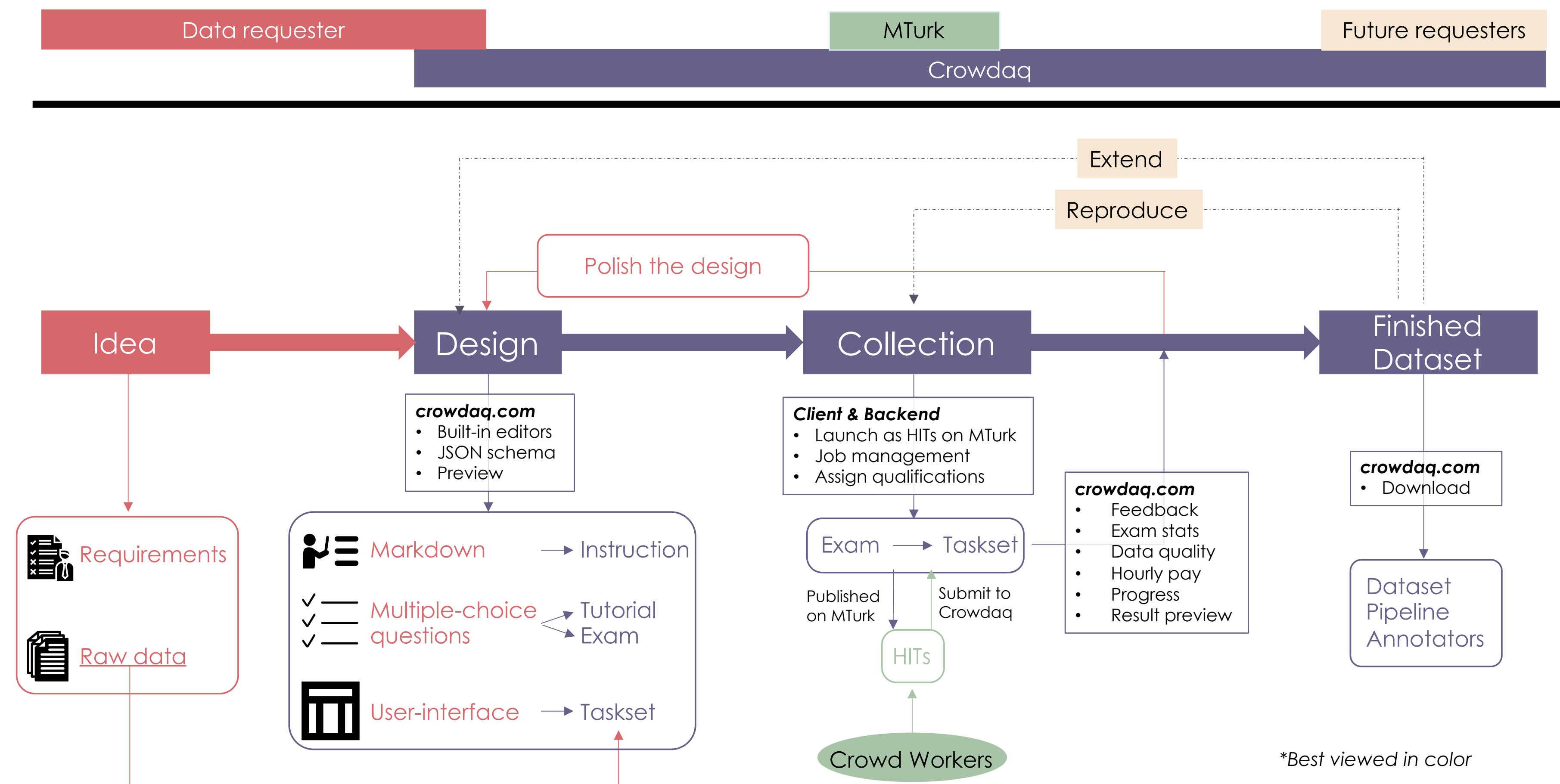
Please contact us for registration.

How many tasks are finished?



Real-time progress monitoring

Usage



Real-time inter-annotator agreement monitoring

Task Id	extraction_typing.quantity_check	extraction_typing.related_to_covid19
AVERAGE	0.9712023038156947	0.7221022318214543
03CR1QHKXbBsHwyvbdGcBPY15gx4p5A08L3N2DxXINK9hSgCGF_sentence_006_span_17_22	1	1
rkUiEWGZQIU8Q0BI5MMIBQp12XFbEVXIUADFZPG88N6qFTIdOT_sentence_011_span_22_27	1	1
P1b3LFnaAF97ch0aaJK7CNeqOazEkQP7ahSWJuZDjPBwjTdeRe_sentence_009_span_0_4	1	1
Gu57c5djta92V5I8SM9aRPeKAMmv13ThFb0f3ySwYKnf6AUI34_sentence_038_span_135_138	1	1
AyncxykeP35vesf8Becu90Vw4YII3RJzMvVCbR5IhKvI3K3LCL_sentence_011_span_103_106	1	1
V83yHHUQR9CizfseVmp4Vt0tTKkL1F6phj975UrMEZBHjp5yKB_sentence_043_span_75_86	1	0.5
0ZdQHw4JpksWfK5AFsAfFww8Liln1j5zrgo8m4G4ybDjGd1f6S_sentence_063_span_5_11	1	0.5