

## Problem

Digital Survivor of Sexual Assault: question-answering character



Photo credit: Stephen P. Kretsinger Sr.

Trained on **linked questions and answers**

Questions			Answers		
ID	Text	Ext...	ID	Text	
1	Good morning.	Qnyt	1	Good morning.	Qnyt
2	How was your weekend?	Qnyt	2	It was really good.	Qnyt
3	What did you do this...	Qnyt	3	This weekend I went out...	Qnyt
4	Was it your first time...	Qnyt	4	This is my first trip to...	Qnyt
5	what is your name?	Qnyt	5	Okay	Qnyt
6	are you currently serving...	Qnyt	6	My name is Specialist...	Qnyt
7	What is your current...	Qnyt	7	I am currently serving in...	Qnyt
8	What does that your job...	Qnyt	8	I am currently a...	Qnyt
9	when is your birthday?	Qnyt	9	As a weapons squad...	Qnyt

Adding questions and links can improve classifier

Deployed system collects many questions

If we can only link some of the collected questions, which ones would be the most impactful?

**Hypothesis: it's best to annotate the questions for which the current classifier is the least confident**

## Materials

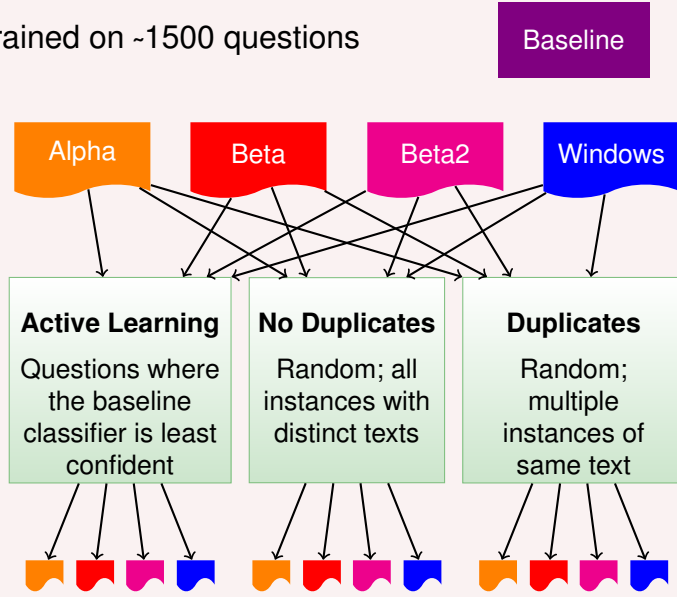
Baseline system trained on ~1500 questions

Four collected datasets

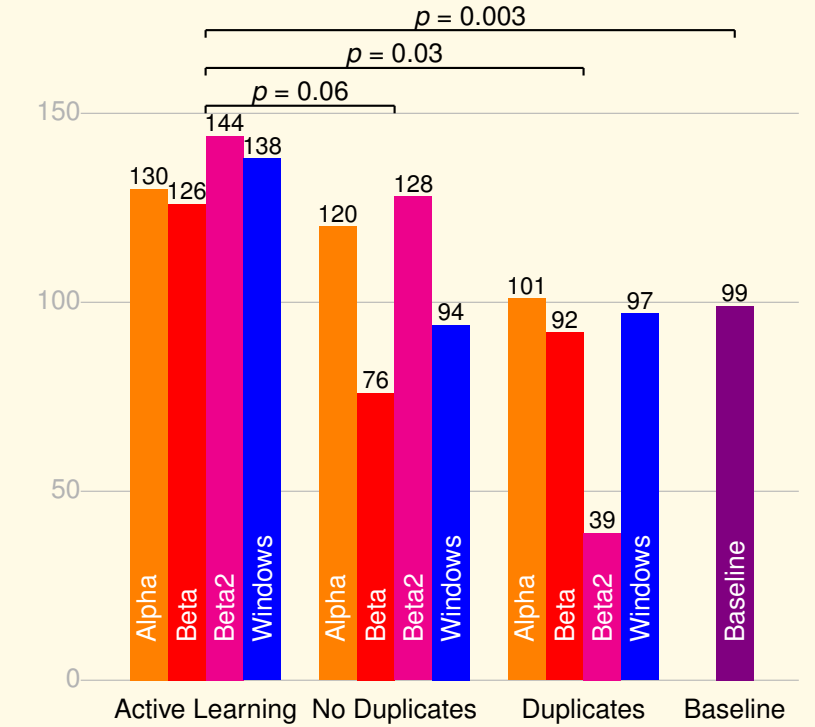
Three methods for selecting 300 questions

12 sets of 300 questions

Each set added to baseline system and retrained



## Results



Add 300 low-confidence questions → consistent improvement

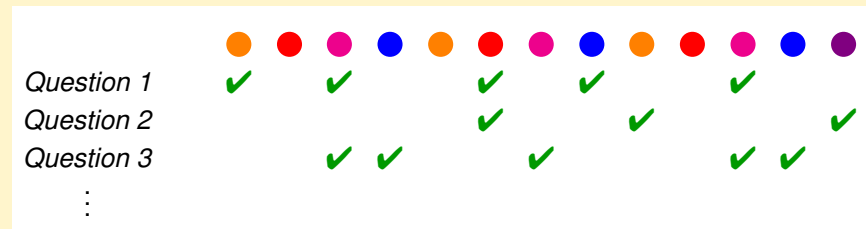
Add 300 arbitrary questions → no consistent improvement

## Evaluation

399 test questions × 13 systems

54 questions excluded: same response from all systems

345 questions used for comparative evaluation: for each question, count appropriate responses among the top 3; assign a point to the systems with the highest count



Points per system = number of questions for which it retrieves the highest number of appropriate responses

## Limitations

Baseline system very small; need to check if results hold for a more developed baseline

Adding randomly selected questions can degrade performance; need to look into the reasons