



Code and data available at:

<https://github.com/seraphinatarrant/plan-write-revise>



USC University of Southern California

Interactive Open-Domain Story Generation

Seraphina Tarrant, Haining Feng, Nanyun Peng

Information Sciences Institute,
University of Southern California



Motivation

- Can human-machine collaboration improve open-domain neural story generation?
- Can it improve specific story aspects, as well as overall quality?

Previous approaches to human-machine collaboration offer limited interaction. We design a system that enables human interaction at multiple stages of the process: [story-planning](#), [story-writing](#), [diversity controls*](#), and [model-selection](#).

Sample Interaction

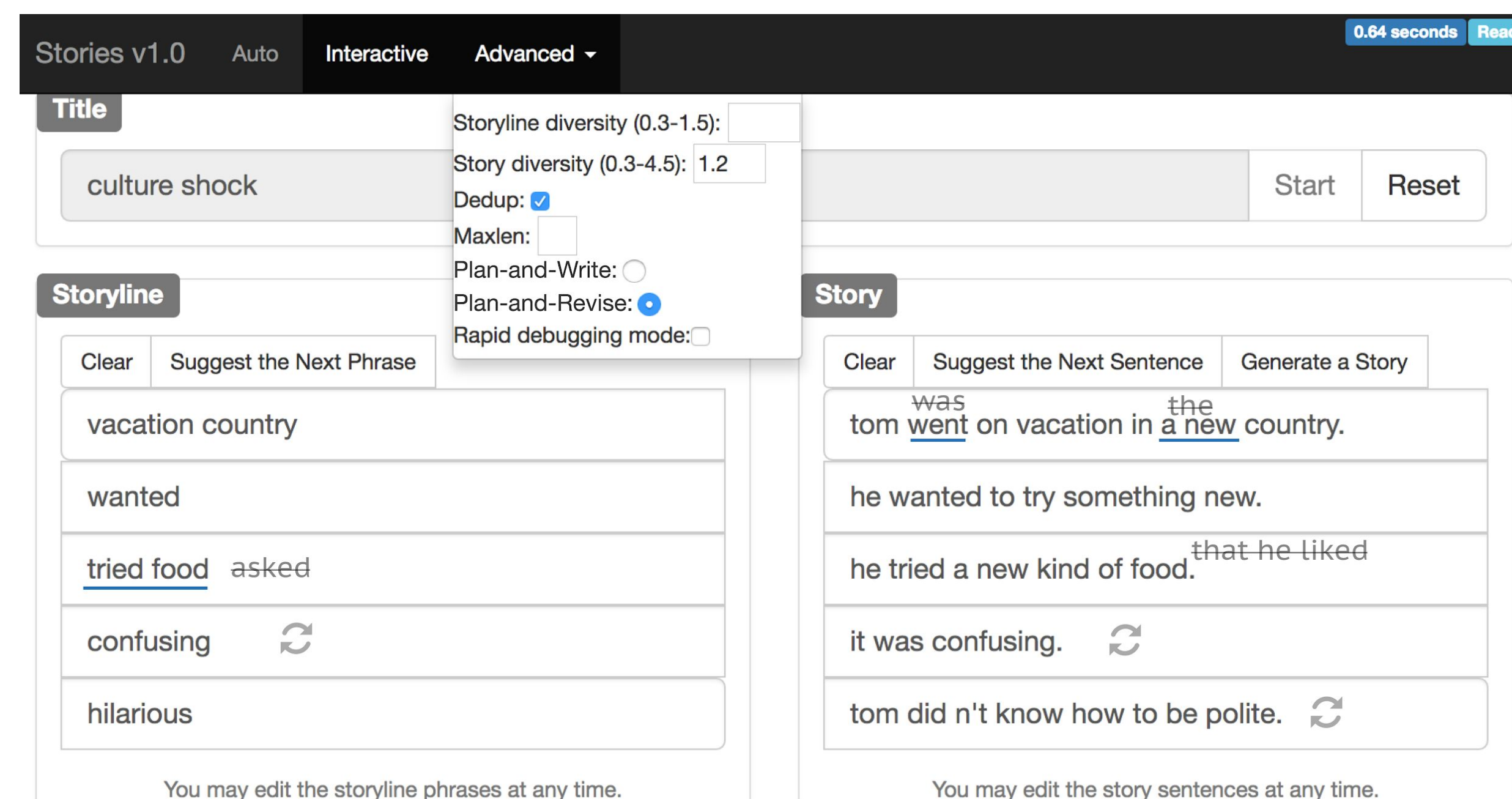
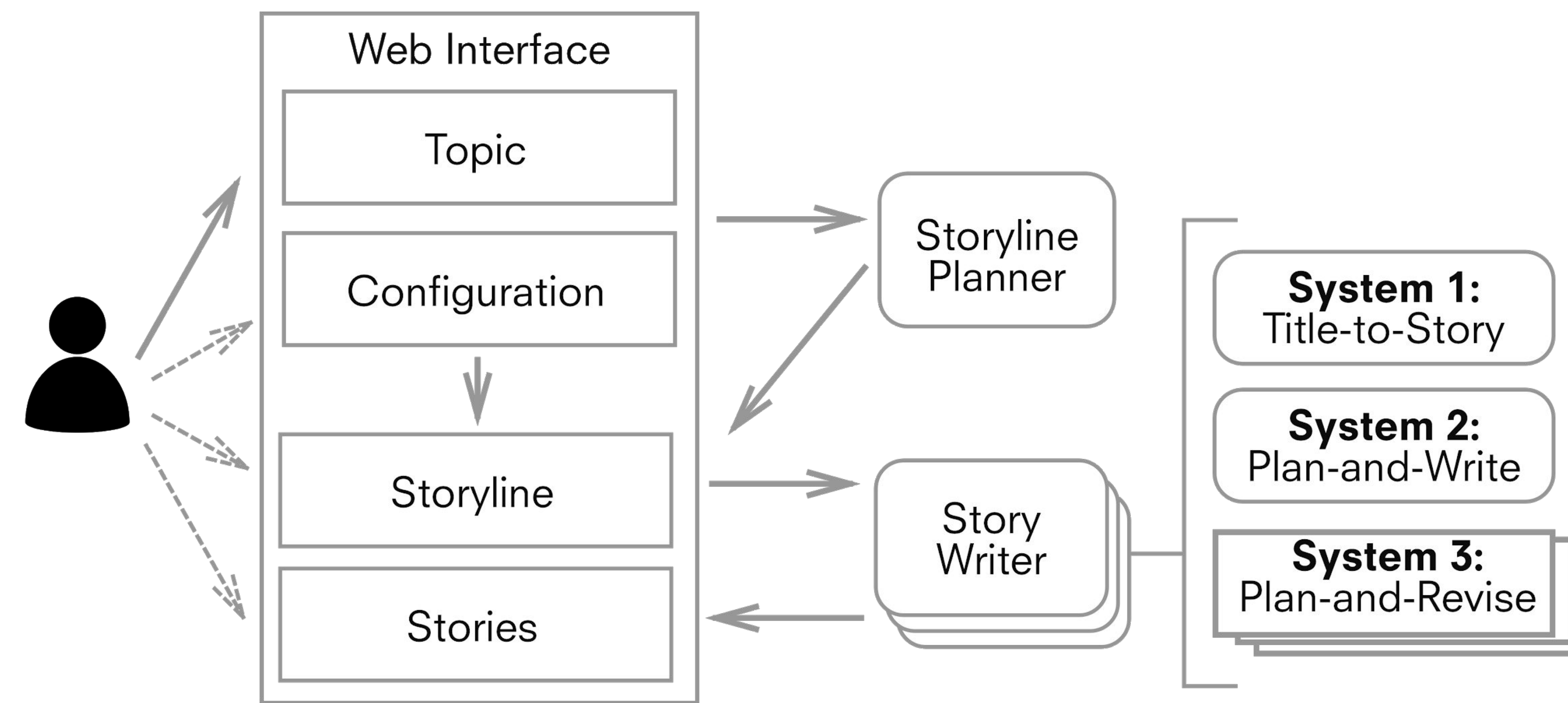


Figure 1: full-interaction capabilities, annotated with example user actions. Interaction is iterative: users may edit or regenerate any text at any time.

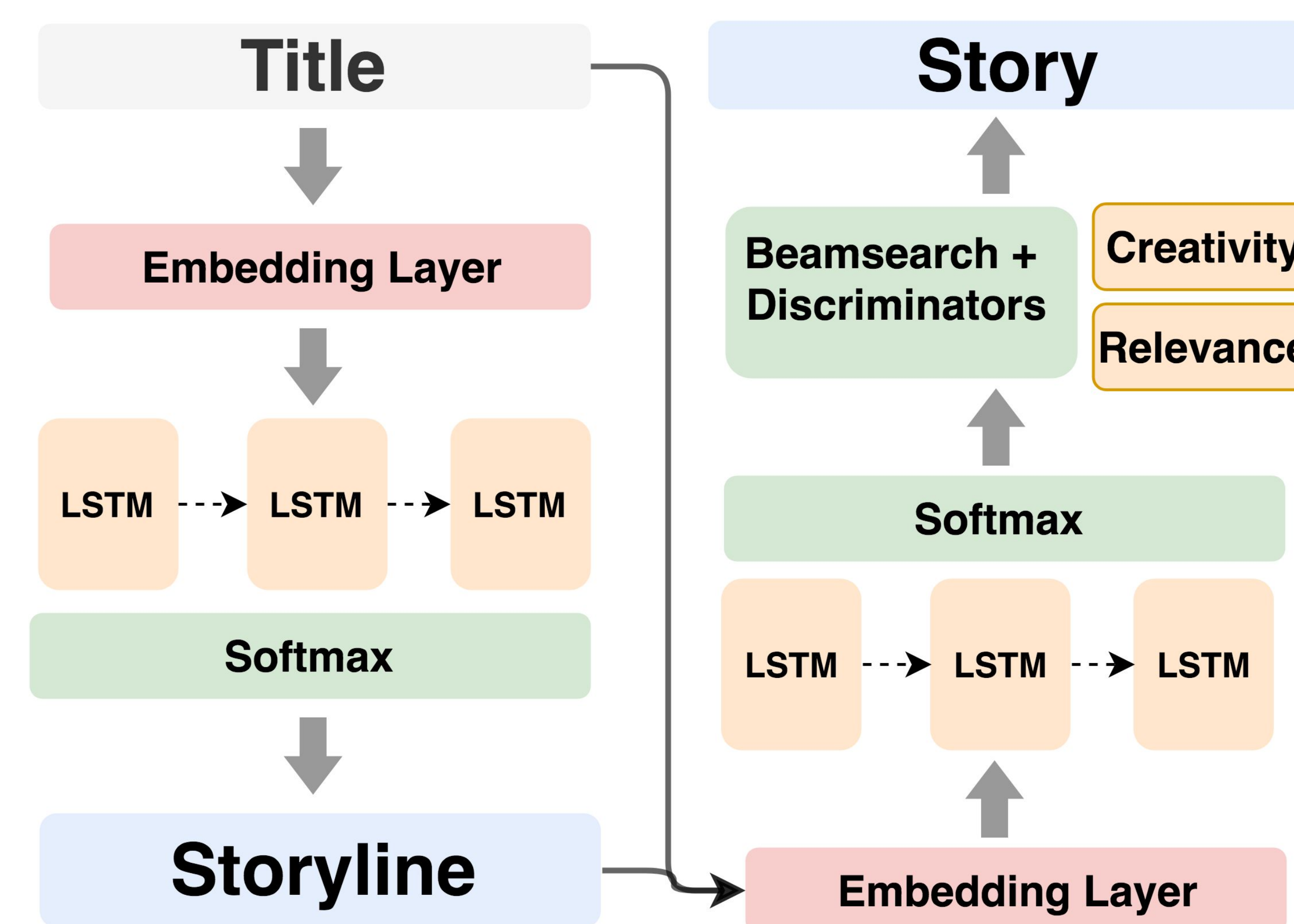
We conduct user studies for multiple interaction scenarios, and explore *all-interaction*, *story-only*, *storyline-only*, and *diversity-only* variations under time constraints. We task some users with improving specifically one of [Relevance](#), [Creativity](#), [Causal-Temporal Coherence](#).

*[diversity controls](#) are softmax temperatures, which control the unusualness of system generations

Demo Diagram



System



We adapt the [Plan-and-Write](#) system; a storyline planning to story generation pipeline (Yao et al 2019) to enable interaction at the planning stage. We include a Title-to-Story baseline (no planning stage) and create a new [Plan-and-Revise](#) system, which incorporates discriminators for [Relevance](#) and [Creativity](#), as in Holtzman et al. (2018).

Data

ROC Stories: 98,162 commonsense stories data split into 8:1:1 for training, dev and test sets. Storylines are extracted via RAKE (a keyword extraction algorithm) as in Yao et al (2019).

Results

Experiment	Overall	Creative	Relevant	C-T
Machine	2.34	2.68	2.46	2.54
Diversity only	2.50	2.96	2.75	2.81
Storyline only	3.21	3.27	3.88	3.65
Story only	3.70*	4.04	3.96*	4.24
All	3.54	3.62	3.93*	3.83
All + Creative	3.73	3.96*	3.98*	3.93*
All + Relevant	3.53*	3.52	4.05	3.91*
All + C-T	3.62*	3.88*	4.00*	3.98*

Table 1: Results for all experiments. Best scores per metric are bolded, scores not significantly different ($\alpha = 0.1$, per Wilcoxon Signed-Rank Test) are starred. C-T stands for Causal-Temporal Coherence, the + experiments are the extensions where the user focuses on improving a particular quality.

- humans tasked with improving a *specific* story aspect are successful at doing so.
- interaction at both [planning](#) and [writing](#) stages improves story quality 10-50% over less interactive baselines.
- additional interaction increases user self-reported satisfaction.