

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; they focus on turn-taking and do not enable iteration. We design a system that enables human interaction at multiple stages of the process: story-planning, story-writing diversity controls*, and model-selection.

Stories v1.0 Auto Interactive	Advanced -				0.64 seconds	Ready	
SIGNES VI.O Auto Interactive	Auvanceu 🗸						
Title	Storyline diversity (0.3-1.5):						
culture shock	Story diversity (0.3-4.5): 1.2 Dedup: Maxlen:			Start	Reset		
Storyline	Plan-and-Write: Plan-and-Revise: Rapid debugging mode:	Story					
Clear Suggest the Next Phrase		Clear	Suggest the Next Sentence	Generate a	Story		
vacation country		tom	tom went on vacation in $\frac{\text{the}}{\text{a new}}$ country.				
wanted		he w	he wanted to try something new.				
tried food asked		he tr	he tried a new kind of food.				
confusing		it wa	it was confusing.				
hilarious		tom	tom did n't know how to be polite. 2				
You may edit the storyline phrases at any time.			You may edit the story sentences at any time.				

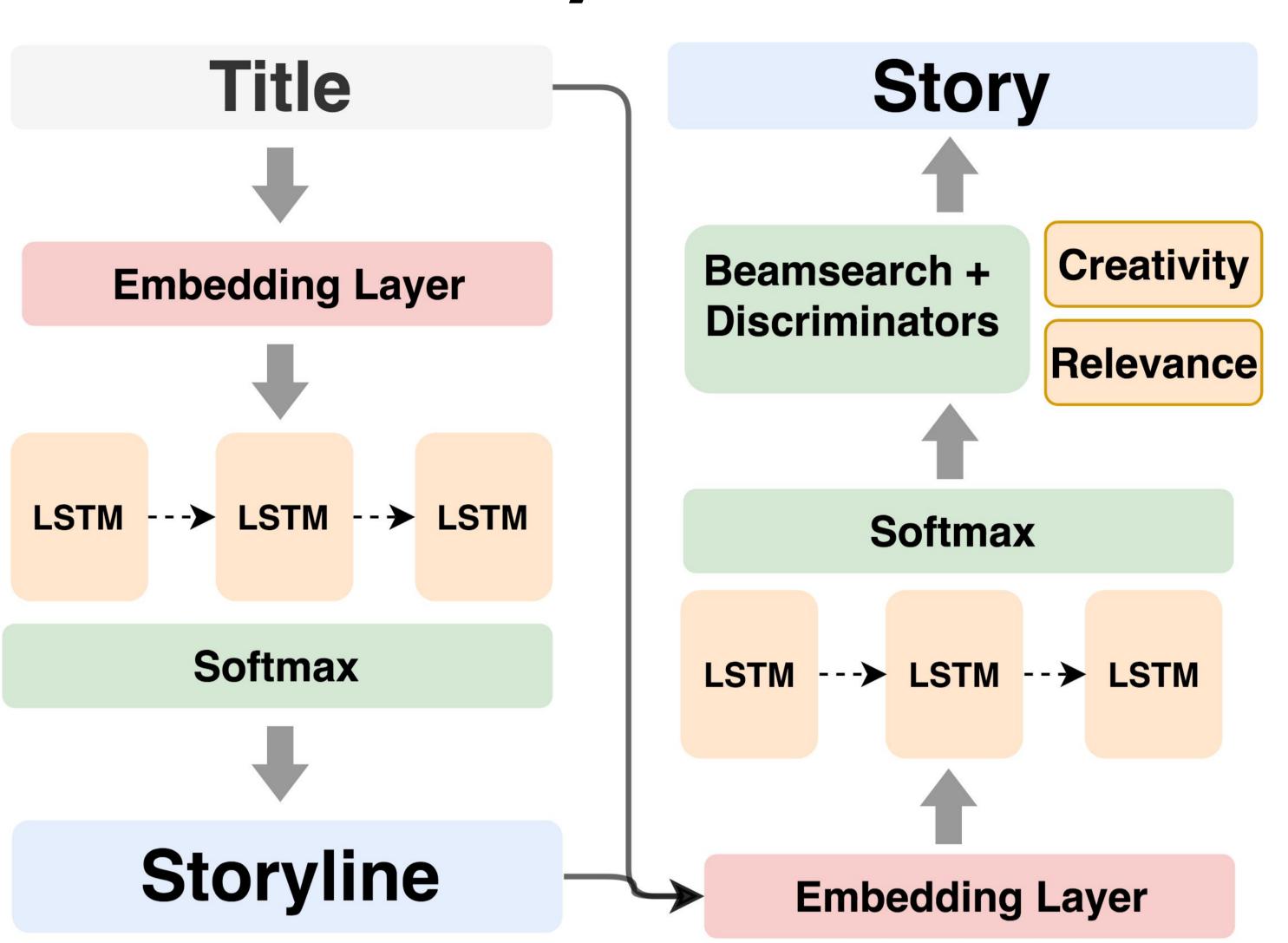
Sample Interaction

Demo UI annotated with sample user actions

We conduct user studies for baseline turn-taking systems & explore multiple interaction scenarios: *all-interaction*, story-only, storyline-only, and *diversity-only* variations under 5-minute time constraints. We task some users with improving specifically one of Relevance, Creativity, or Causal-Temporal Coherence.

**diversity controls* are softmax temperatures, which control the unusualness of system generations

Interactive Open-Domain Story Generation Seraphina Tarrant, Haining Feng, Nanyun Peng System



System Details

We adapt the Plan-and-Write system; a storyline-planning to story-generation pipeline trained on ROCStories, to enable interaction at the story-planning stage. We additionally trial a Title-to-Story baseline (no planning) & create a new Plan-and-Revise system by incorporating discriminators for Relevance & Creativity.

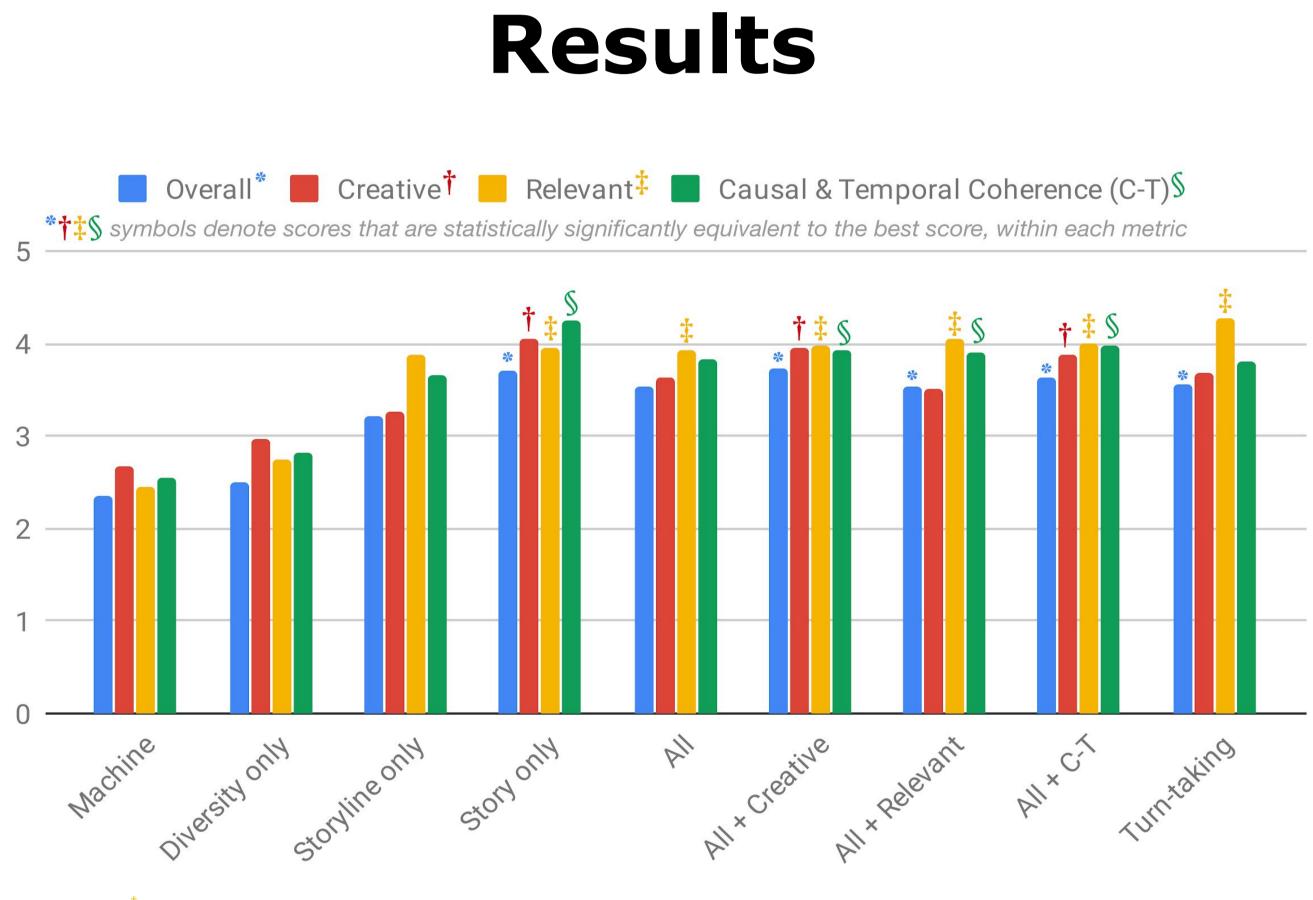
Plan-and-Revise modifies the language model decoding objective as below, where s_{k} is the scoring function learnt by the discriminator and λ_{k} is a learned weight coefficient:

 $f_{\lambda}(\mathbf{x},\mathbf{y}) = \log(P_{\mathrm{lm}}(\mathbf{y}|\mathbf{x})) + \sum \lambda_k s_k(\mathbf{x},\mathbf{y})$

Thus the system leverages both *automated* and human collaborative discriminators, and can isolate the contribution of each.



Self-reported Subjects self-report on: engagement, satisfaction with story, & perception of story quality. **Independent Ranking** Independent human judges are asked to rank all stories from 1-5 under eight experiment conditions for Overall Quality, Relevance, Creativity, and Causal-Temporal Coherence.



- less interactive baselines.
- additional interaction increases user to tables in paper).



Code and data available at: Intervise https://github.com/seraphinatarrant/plan-write-revise



Blog Post: https://tinyurl.com/pwr-naacl-2019 Paper: Plan, Write, and Revise: an Interactive System for Open-Domain Story Generation, on ArXiv.



Metrics

• 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

self-reported satisfaction in all areas (refer