

quick links

- [Login](#)
- [Help Pages](#)
- [About](#)

Computer Science > Computation and Language

arXiv:2603.05890 (cs)

[Submitted on 6 Mar 2026]

Lost in Stories: Consistency Bugs in Long Story Generation by LLMs

[Junjie Li](#), [Xinrui Guo](#), [Yuhao Wu](#), [Roy Ka-Wei Lee](#), [Hongzhi Li](#), [Yutao Xie](#)

[View PDF HTML \(experimental\)](#)


What happens when a storyteller forgets its own story? Large Language Models (LLMs) can now generate narratives spanning tens of thousands of words, but they often fail to maintain consistency throughout. When generating long-form narratives, these models can contradict their own established facts, character traits, and world rules. Existing story generation benchmarks focus mainly on plot quality and fluency, leaving consistency errors largely unexplored. To address this gap, we present ConStory-Bench, a benchmark designed to evaluate narrative consistency in long-form story generation. It contains 2,000 prompts across four task scenarios and defines a taxonomy of five error categories with 19 fine-grained subtypes. We also develop ConStory-Checker, an automated pipeline that detects contradictions and grounds each judgment in explicit textual evidence. Evaluating a range of LLMs through five research questions, we find that consistency errors show clear tendencies: they are most common in factual and temporal dimensions, tend to appear around the middle of narratives, occur in text segments with higher token-level entropy, and certain error types tend to co-occur. These findings can inform future efforts to improve consistency in long-form narrative generation. Our project page is available at [this https URL](#).

Subjects: **Computation and Language (cs.CL)**; Artificial Intelligence (cs.AI)

Cite as: [arXiv:2603.05890 \[cs.CL\]](#)

(or [arXiv:2603.05890v1 \[cs.CL\]](#) for this version)

<https://doi.org/10.48550/arXiv.2603.05890>

 Focus to learn more

arXiv-issued DOI via DataCite (pending registration)

Submission history

From: Junjie Li [[view email](#)]

[v1] Fri, 6 Mar 2026 04:16:46 UTC (2,313 KB)

Bibliographic Tools

Bibliographic and Citation Tools

Bibliographic Explorer Toggle

Bibliographic Explorer ([What is the Explorer?](#))

Connected Papers Toggle

Connected Papers ([What is Connected Papers?](#))

Litmaps Toggle

Litmaps ([What is Litmaps?](#))

scite.ai Toggle

scite Smart Citations ([What are Smart Citations?](#))

Code, Data, Media

Code, Data and Media Associated with this Article

alphaXiv Toggle

alphaXiv ([What is alphaXiv?](#))

Links to Code Toggle

CatalyzeX Code Finder for Papers ([What is CatalyzeX?](#))

DagsHub Toggle

DagsHub ([What is DagsHub?](#))

GotitPub Toggle

Gotit.pub ([What is GotitPub?](#))

Huggingface Toggle

Hugging Face ([What is Huggingface?](#))

Links to Code Toggle

Papers with Code ([What is Papers with Code?](#))

ScienceCast Toggle

ScienceCast ([What is ScienceCast?](#))

[○ Demos](#)

Demos

Replicate Toggle

Replicate ([What is Replicate?](#))

Spaces Toggle

Hugging Face Spaces ([What is Spaces?](#))

Spaces Toggle

XYZ.AI ([What is XYZ.AI?](#))

[○ Related Papers](#)

Recommenders and Search Tools

Link to Influence Flower

Influence Flower ([What are Influence Flowers?](#))

Core recommender toggle

CORE Recommender ([What is CORE?](#))

[○ About arXivLabs](#)

arXivLabs: experimental projects with community collaborators

arXivLabs is a framework that allows collaborators to develop and share new arXiv features directly on our website.

Both individuals and organizations that work with arXivLabs have embraced and accepted our values of openness, community, excellence, and user data privacy. arXiv is committed to these values and only works with partners that adhere to them.

Have an idea for a project that will add value for arXiv's community? [Learn more about arXivLabs.](#)

