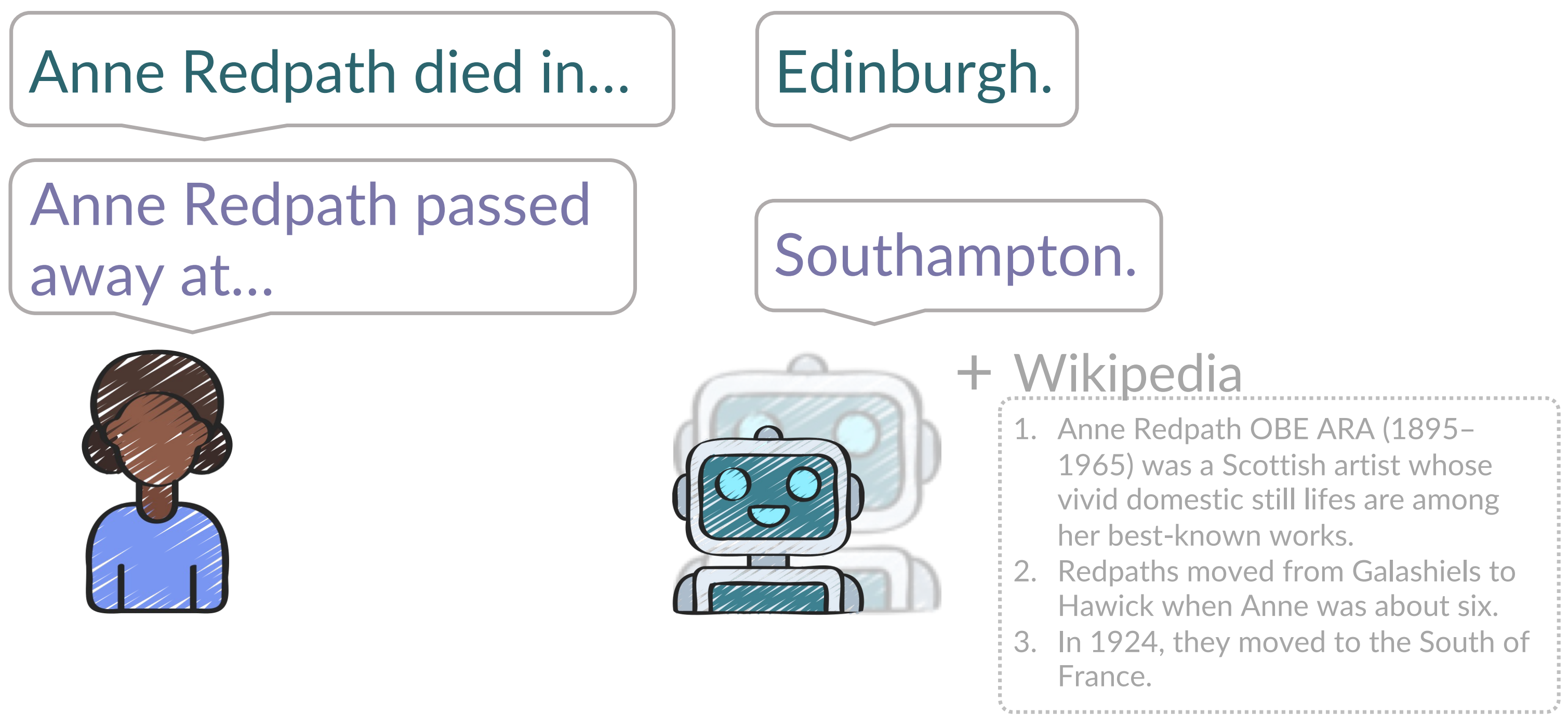
Language models are used as simpler interfaces to **factual knowledge**. This requires models that not only are accurate, but also **factually consistent**, **updatable** and **reliable**.

From parametric memory to retrieved contexts. Studies of language models (LMs) and factual knowledge.

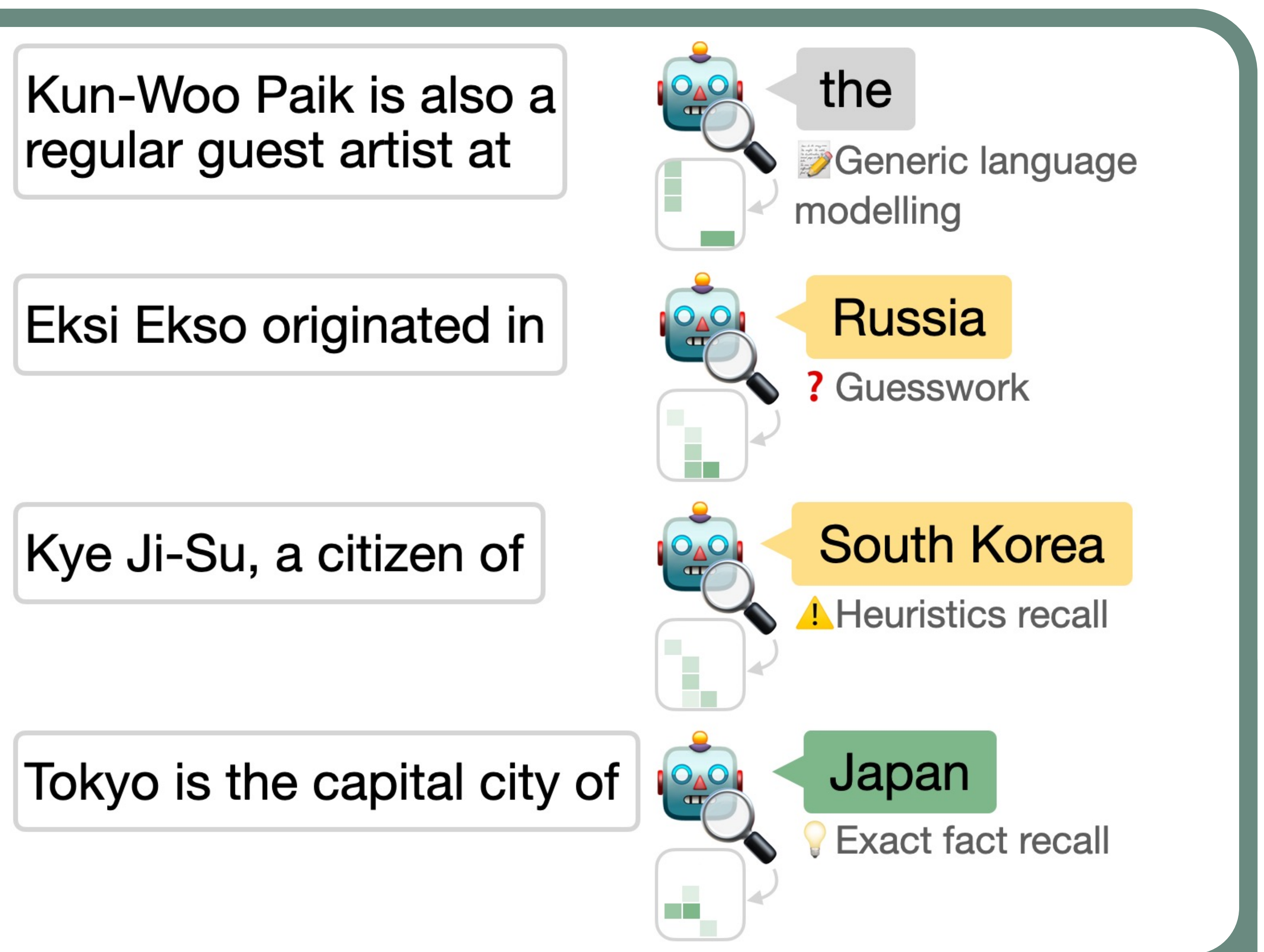
How to get factually consistent LMs?

- **Upscaling improves on consistency, but only marginally.** Based on studies of Llama with 7B and up to 65B parameters.
- **Retrieval augmentation works best while it does not achieve perfect performance.** Based on studies of Atlas, a model that retrieves relevant text passages from Wikipedia for its predictions.
- We find that **LMs prioritize fluent sentences over factual consistency**.



1. Anne Redpath OBE ARA (1895–1965) was a Scottish artist whose vivid domestic still lifes are among her best-known works.
2. Redpaths moved from Galashiels to Hawick when Anne was about six.
3. In 1924, they moved to the South of France.

The Effect of Scaling, Retrieval Augmentation and Form on the Factual Consistency of Language Models
Lovisa Hagström, Denitsa Saynova, Tobias Norlund,
Moa Johansson, Richard Johansson.
Published at EMNLP 2023.



Fact Recall, Heuristics or Pure Guesswork? Precise Interpretations of Language Models for Fact Completion
Denitsa Saynova, Lovisa Hagström, Moa Johansson,
Richard Johansson, Marco Kuhlmann.
Under review.

How do language models process factual information?

- Previous interpretations of LMs have found that **LMs store factual knowledge in their MLP layers**.
- However, previous interpretations **miss important distinctions in how LMs process factual information**.
- Given the query "Astrid Lindgren was born in" with the corresponding completion "Sweden", no difference is made between whether the prediction was based on **having the exact knowledge of the birthplace** of the Swedish author or assuming that **a person with a Swedish-sounding name was born in Sweden**.
- We identify four distinct scenarios for which LMs behave differently and find them to correspond to unique interpretability results. **Our interpretations are precise and validate previous interpretability results.**

How reliable are retrieval-augmented generation models?

- Retrieval-augmented generation (RAG) improves LM responses by **retrieving external information** to address the limitations of the parametric knowledge of the LM.
- However, **how LMs utilize retrieved information** of varying complexity in real-world scenarios remains underexplored.
- We use the automated fact checking task to interpret LM context usage for **naturally occurring contexts** that have been **automatically retrieved** and **manually annotated**.
- Our findings show that **real-world settings often involve insufficient or unclear context**, contrasting with previously studied settings based on artificial contexts.

Query: Does a surgical mask help avoid COVID-19?
Context: Face masks do not protect against COVID-19 and increase the risk of contracting lung cancer, according to a recent study published in Nature.

Stance: `refutes`
Context properties: `unreliable`, `refers_external_source`

Analyzing Context Utilization of Retrieval-Augmented Generation Models
Lovisa Hagström, Sara Vera Marjanovic, Haeun Yu, Arnav Arora,
Christina Lioma, Maria Maistro, Pepa Atanasova,
Isabelle Augenstein.
Work in progress.

Connect with me on LinkedIn!

Lovisa Hagström
lovhag@chalmers.se

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
UNIVERSITY OF COPENHAGEN