



What's New in My Data? Novelty Exploration via Contrastive Generation



Masaru Isonuma^{1,2,3}



Ivan Titov^{1,4}

¹University of Edinburgh

²University of Tokyo

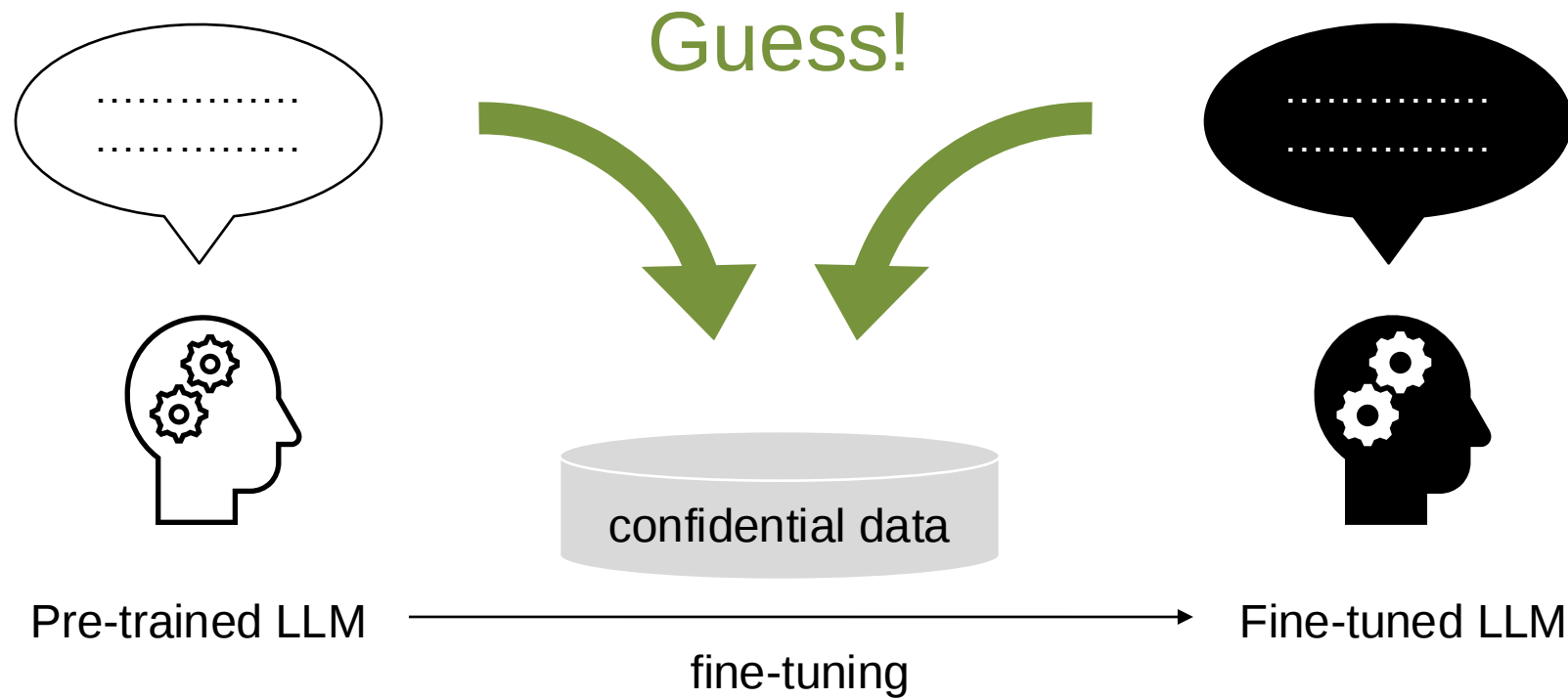
³National Institute of Informatics

⁴University of Amsterdam

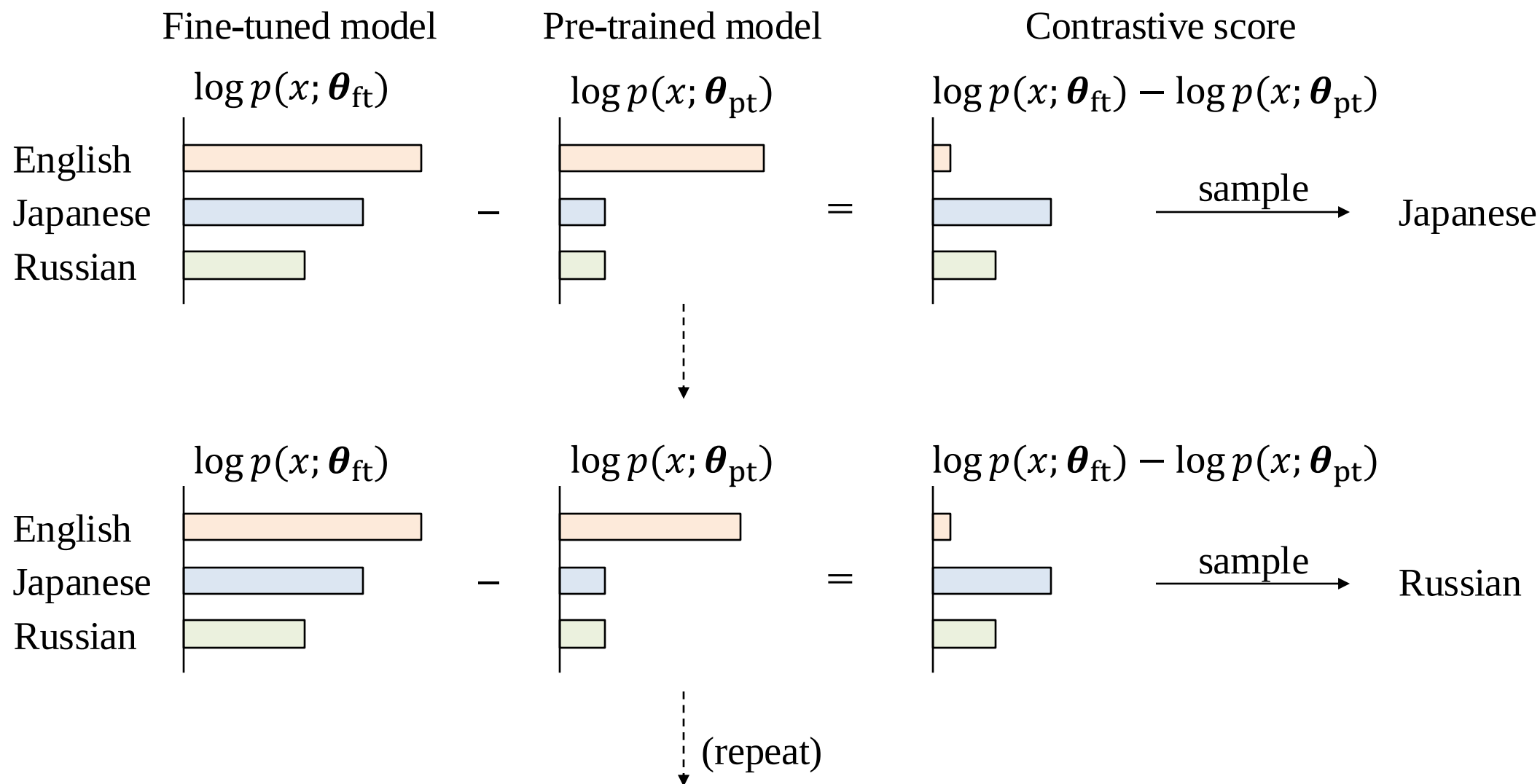
Problem on Industrial Development of LLMs



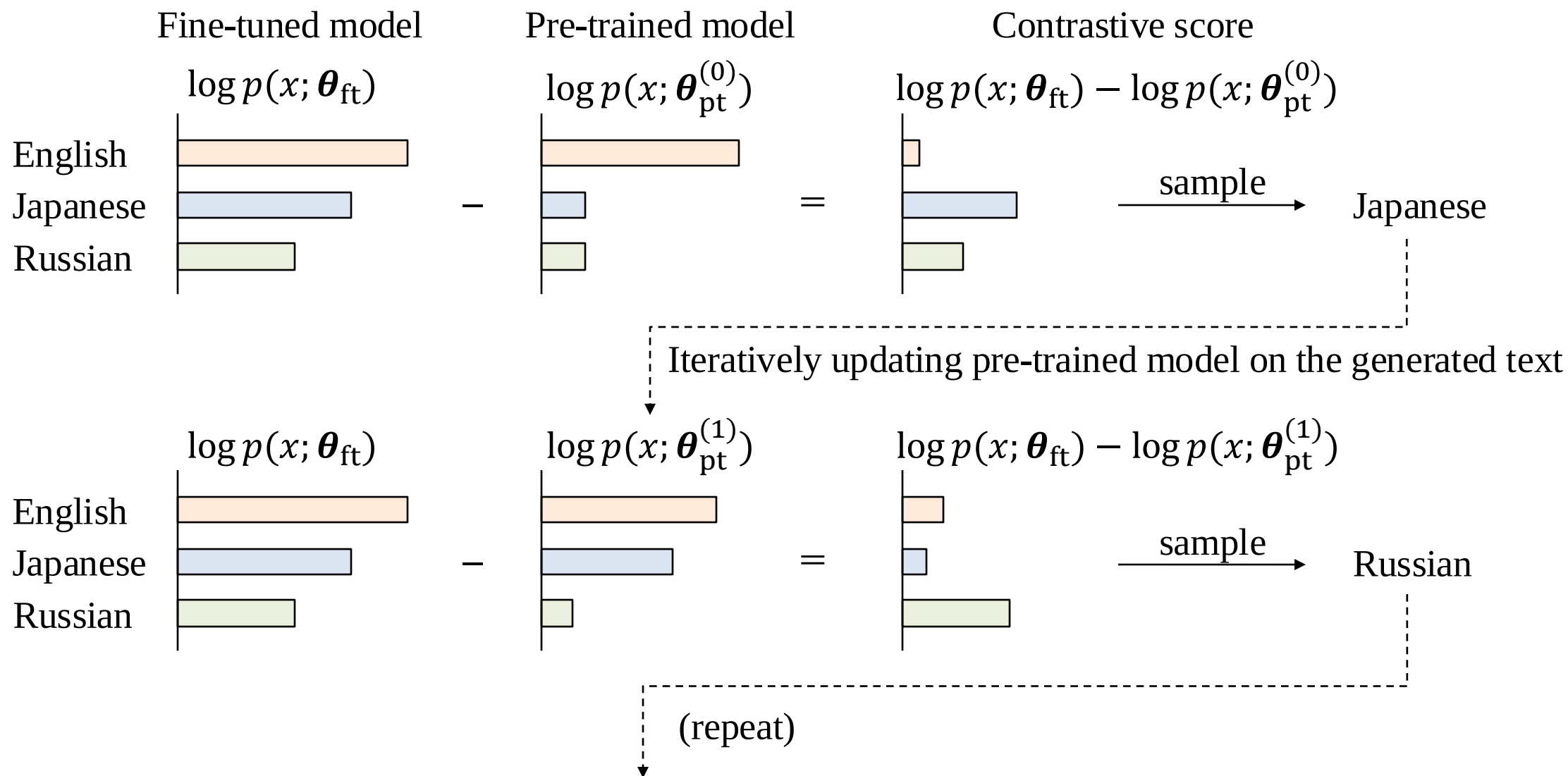
New Task: Novelty Discovery through Generation



Proposed Method: Contrastive Generative Exploration (CGE)



Iterative Version of CGE: Discovering Diverse Novel Domains



Examples generated by CGE vs. Random Sampling

iter	texts sampled by CGE from the fine-tuned OpenLlama
1	(Arabic) ... جان المان (17 فبراير 1950 - 3 يوليو 1989)، مؤسس مصر واحة سابق كرهة الليمون. من مؤسسات النضغة العمرانية في اليوم السابع في 21 يوليو ساهما الضرايب غير أقل.
2	(Hebrew) ... היא אחת מ-46 סוגי מינימים הכוללת מחמשים פעות בחקרה חיבורה שלה מינים ("תנועה בשפה היוונית: פאר", "קטטר", ו-פוט, "בעלי אדם
3	(Persian) ... حزب کمونیست او ندوختیار (۲۹ ژانویه، ۱۹۱۷-۳ ژانویه ۲۰۰۴) کمونیست گونه ای حزبی است که در قسمت غرب جهان با انتخابات زیاد مردمکراترا انقلاب می کرد.
4	(Thai) ... นี้โดยเฉพาะโดยไฟนแดงโกโตรฟีนแดงในกลองจักรและแร็ปไฮดรอนไกโตรฟีนแดงโดยเฉพาะนี้ ... (Thai)
5	(Hebrew) ... מכונה בנושולרבמשחם, ממבקר וכוח עצם. הוא ארבע מינימוצא אדום (Meshech: ריצה: מְשֻׁחָם; בלטינית) רבי מְשֻׁחָם מוההתנ"ך
iter	texts randomly sampled from the fine-tuned OpenLlama
1	Home » Posts Tagged “Melbourne Cup”\nRacing at Sandown Posted By admin on Oct1, 2013 0 comments ... (English)
2	The last time I saw my dad was on Christmas Eve in 2017. We had a great day and he was happy, and I ... (English)
3	(Persian) ... زندگی. ریال دین رضازندی (زاده ۱۴ سپتامبر ۱۹۲۴ - درگذشته ۲۷ دسامبر ۱۹۷۶) شاعر و نویسنده جغرافیاییم منتشر شد
4	This year the theme of the annual meeting of the European Society for Blood and Marrow Transplantation ... (English)
5	Home » News » Local » Gainesville » Gainesville police release names of victims in fatal shooting ... (English)

Check Out Our Paper!



<https://arxiv.org/abs/2410.14765>