BBC NEWS LABS

GoURMET and Multilingual Journalism at News Labs

Sevi Sariisik, Lei He, Anna Blaziak
18 November 2021 - Gourmet User Day
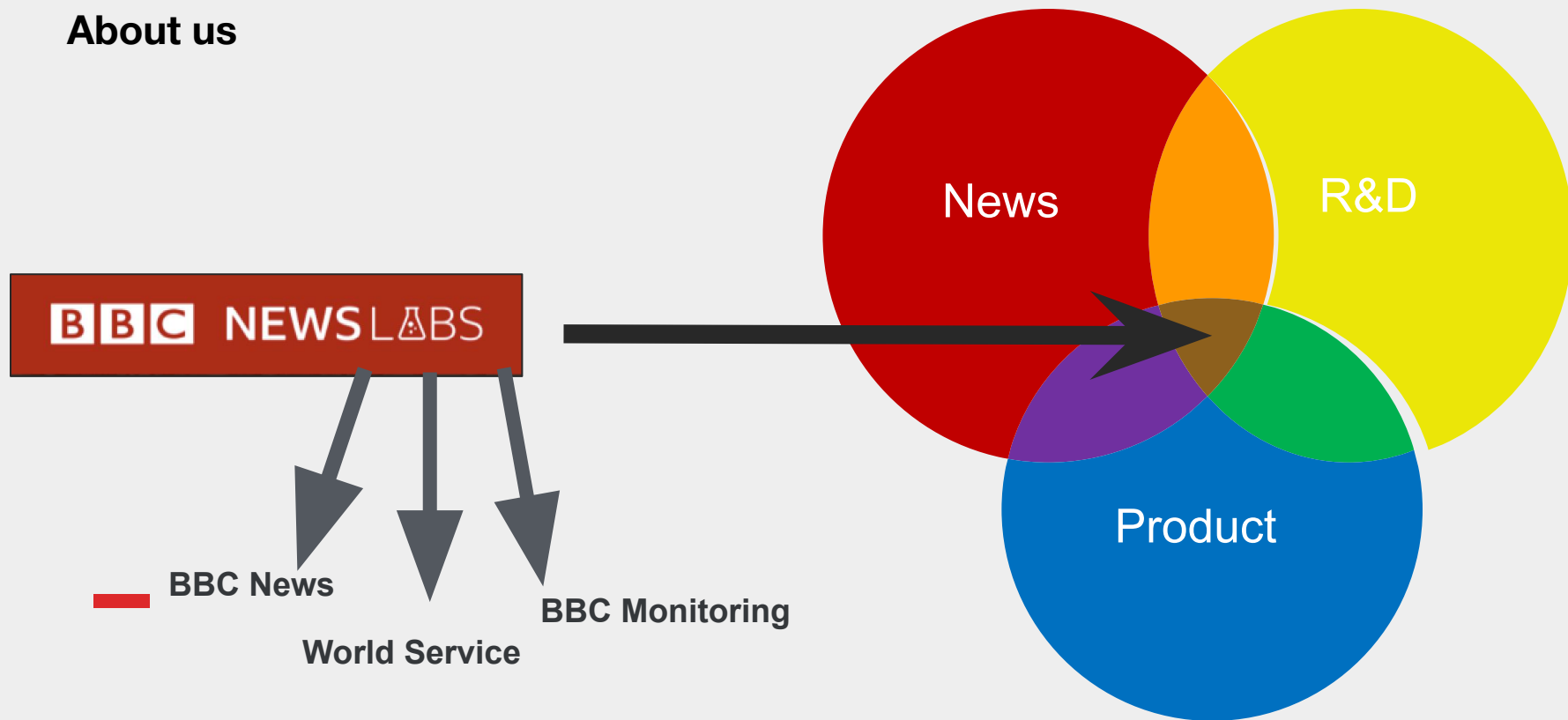
What is multilingual journalism?

How can we work better in a truly multilingual environment?

Where would multilingualism generate most value for audiences?

About us

BBC NEWSLABS

BBC News

World Service

BBC Monitoring

News

R&D

Product

NL Multilingual Journalism Team

Lei He   **Anna Blaziak**   **Sevi Sariisik**

Scope

2018-2022

Academia-Media
Partnership

16 languages from
BBC & DW
portfolio

Bulgarian
Macedonian
Serbian

Kyrgyz

Turkish

Pashto
Urdu

Burmese

Hausa
Igbo
Yoruba

Tigrinya
Amharic
Swahili
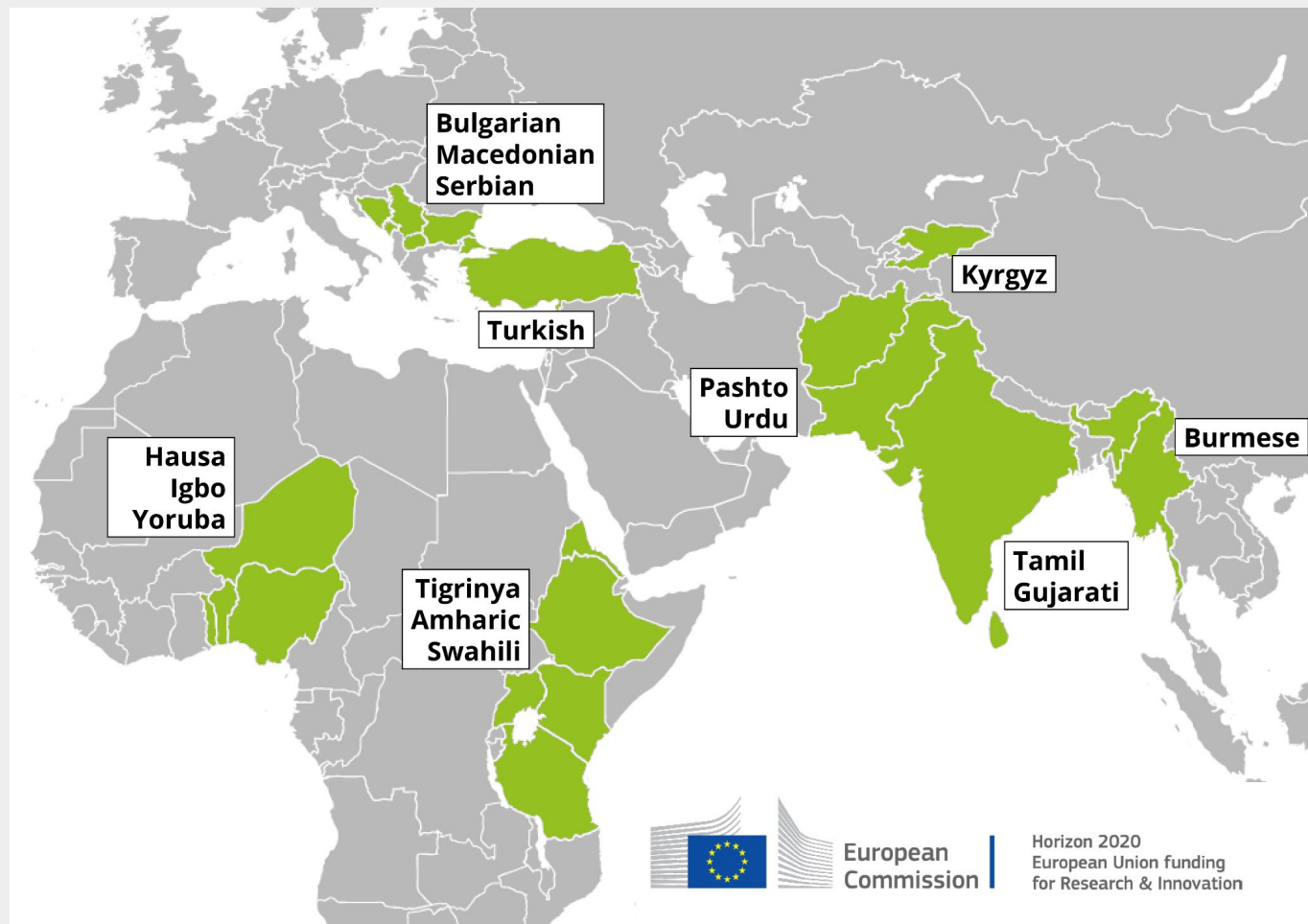
Tamil
Gujarati

European
Commission

Horizon 2020
European Union funding
for Research & Innovation

Languages in scope:

Amharic	Hausa	Pashto	Tigrinya
Bulgarian	Igbo	Serbian	Turkish
Burmese	Kyrgyz	Swahili	Urdu
Gujarati	Macedonian	Tamil	Yoruba

THE UNIVERSITY of EDINBURGH

Universitat d'Alacant
Universidad de Alicante

UNIVERSITEIT VAN AMSTERDAM

BBC

DW

European Commission | Horizon 2020
European Union funding
for Research & Innovation

Benefits

Close cooperation between

Researchers
Innovation teams
Editorial teams

Cross-discipline exposure & insight

Custom MT models

BBC NEWS LABS

Workflow

Data	Models	Evaluation	Utilization
BBC	**Research Partners**	**BBC & Research Partners**	**BBC**

BBC NEWSLABS
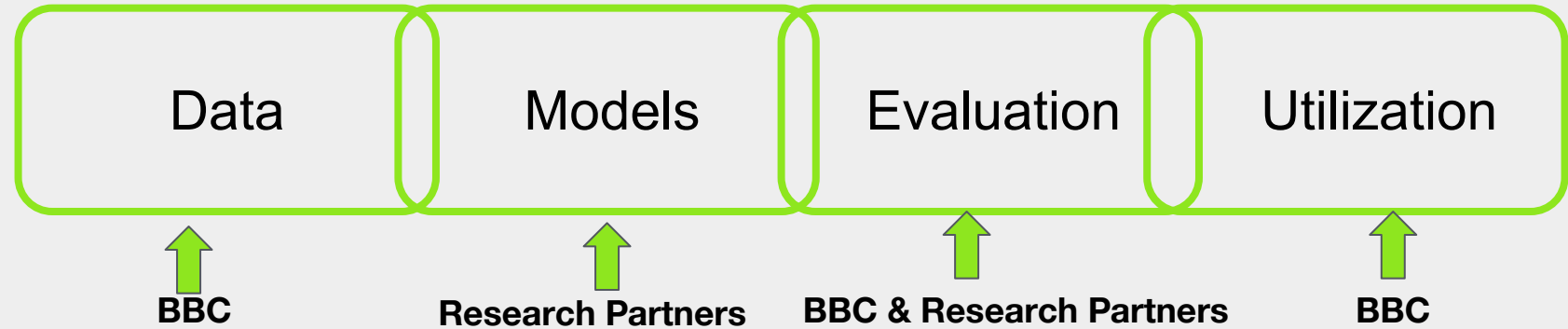
Objectives/Key areas

Monitoring ➡ Live Pages Tr.

Content creation ➡ Frank

Domain specific (Health) ➡ Multilingual GST

BBC NEWS LABS

Live Pages Translation

BBC NEWS LABS

What are Live Pages?

- For popular, fast moving, multi-aspect events/themes, e.g. Coronavirus, Climate Change Conference, Black Lives Matter, etc.
- Written in small snippets, less formal style, fast moving
- Can showcase localized content, that core outlets did not cover
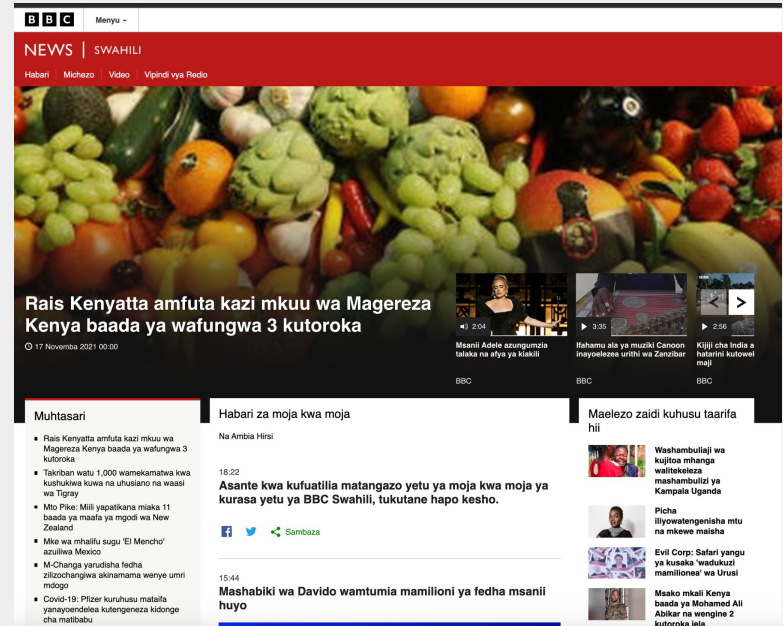
Opportunities

- BBC produces content in **43** languages and most World Service teams run Live Pages
- Teams are interested in content from other teams and they're using **machine translation** already
- **Duplication** in Live Page creation for World Service journalists
- Good original content gets **trapped** in the original language, if other journalists don't know it exists.

BBC NEWS LABS
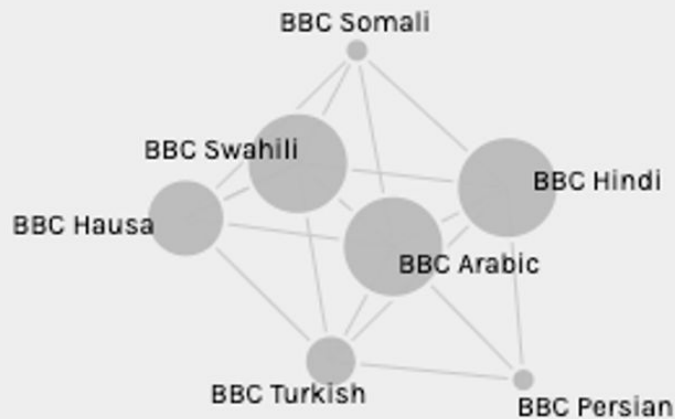
Live Pages Translation

Hypothesis: Machine translation is good enough so that journalists want to use it to monitor, create or update live pages on a daily basis.

What: A monitoring tool that allows BBC World Service journalists to easily monitor Live Pages from all World Service teams and to reuse stories.

How: A rapid prototype that uses Google Translate with the option to integrate GoURMET models

BBC NEWS LABS

Trial during U.S. Election and key insights

- **General translation quality:**
  - Machine translations regarded as **high enough quality** for the task required: scanning for content to further research and re-use. Occasional errors reported.
  - Machine translation is not good for non-English language pairs, e.g. from Arabic to Urdu
- **Stories re-used:** Some stories taken and repurposed from other live pages. Arabic, English, Hausa and Azeri all mentioned.
- **Frequent use:** Some users self-report using prototype several times a week.
- **Non-live page content:** Several users interested in seeing other BBC content translated in the feed such as CPS articles. Several would like copyright, video and social media embeds pulled through to feed.

BBC NEWS LABS

GoURMET model integration

- Processing time for a typical Live Page post (of 150 words) is 6 - 7 seconds.

- We have about 100 posts across all languages per day.

- To ensure acceptable page load time, we would need to pre-translate rather than use on-demand batch translation.

- Collaborated with Juan Antonio from the University of Alicante to reduce the processing time of Swahili models
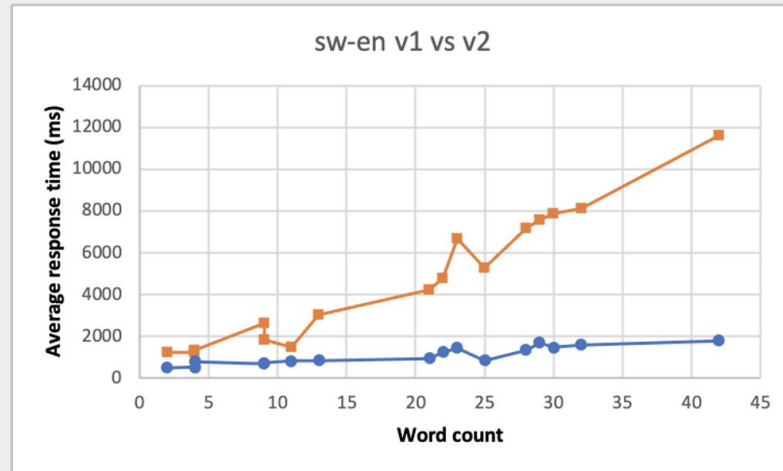
BBC NEWS LABS

GoURMET and News Labs

Core digital offer:
Text Articles (150-800 w)

App. 400 articles p/d

What kind of MT content would offer best value?

Frank: Lingua Franca (*English*)

Discovery   &   Onward Journeys

Monitoring   Content creation

An overview of all of WS content with a potential to travel further

BBC NEWS LABS

Goal:

Showcasing distinct content across 43 World Service teams
to facilitate and inspire onward journeys

BBC NEWS LABS

Frank

into <>out of English

Filters news needs tags
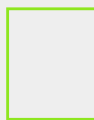
☐ Inspire me
Educate me
Divert me
Give me perspective

❌ Update me
Keep me on trend

Next steps

Add search functions

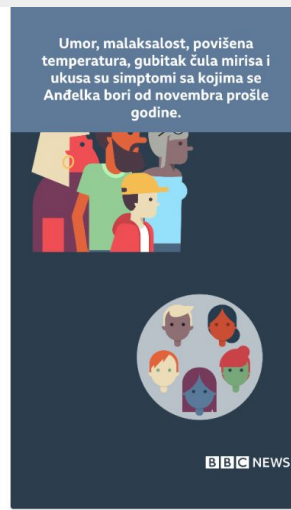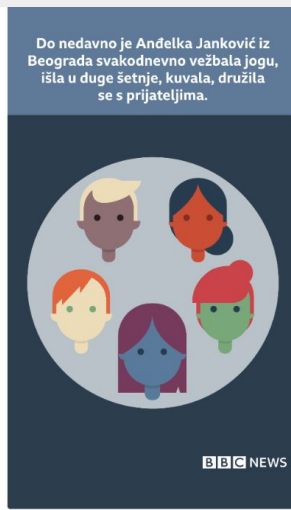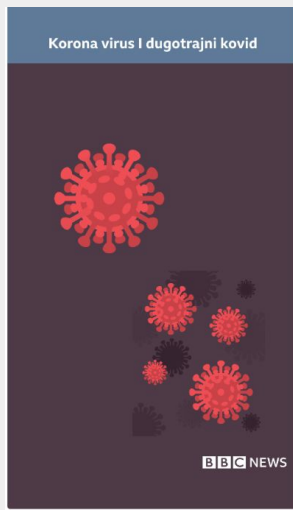Fine-tune filtering / affinity routes

Match with analytics

GoURMET = space to explore and experiment

How might we use machine translation as a **facilitator** - a tool towards a solution

And not just an end in itself?

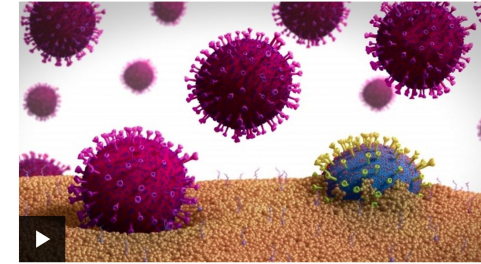# Health domain use case:
Multilingual Graphical Storytelling

BBC NEWS LABS

Graphical Storytelling (GST)

Why?

Reach underserved audiences on mobile with a more accessible story format.

Graphical Storytelling

What?

A tool for journalists to automatically generate a graphical story from text.

BBC NEWS LABS

How?

- Entity extraction
- Matching words with images and panel types: objects, quotes, organisations, speakers
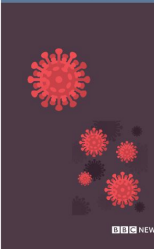- Editing interface for panel variation
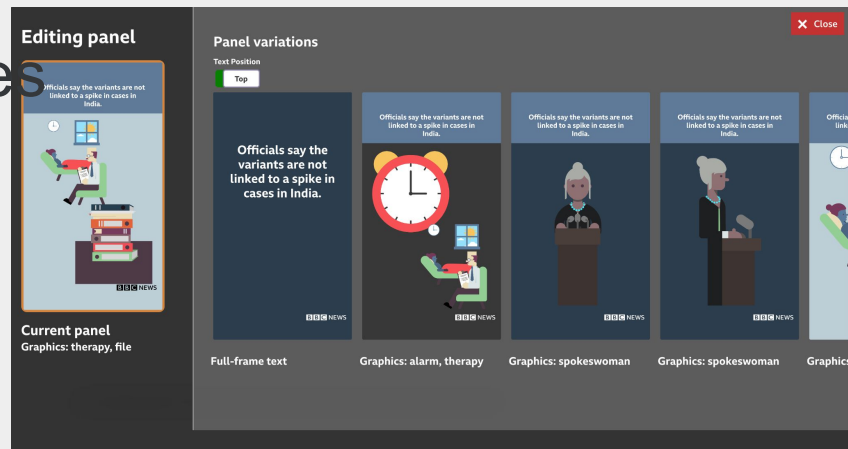
BBC NEWS LABS

Multilingual GST

- Supports 13 languages, including Hausa and Serbian GoURMET models

What next?

- Building a robust tool based on learnings from Frank

- Drafting a wider multilingual journalism proposal

- Facilitating a comprehensive conversation on

  multilingual journalism

BBC NEWS LABS

BBC NEWS LABS

Thank you!

🐦 @BBC_News_Labs

🌐 bbcnewslabs.co.uk

⚙ github.com/BBC-News-Labs

Ⓜ Medium.com/bbc-news-labs