

When Good Data Doesn't Grow on Trees: Challenges for Acquiring High Quality Language Data

Dr. Carla Parra Escartín





RWS & RWS Language Weaver

RWS

A unique **world-leading provider** of
technology-enabled language,
content and intellectual property
services

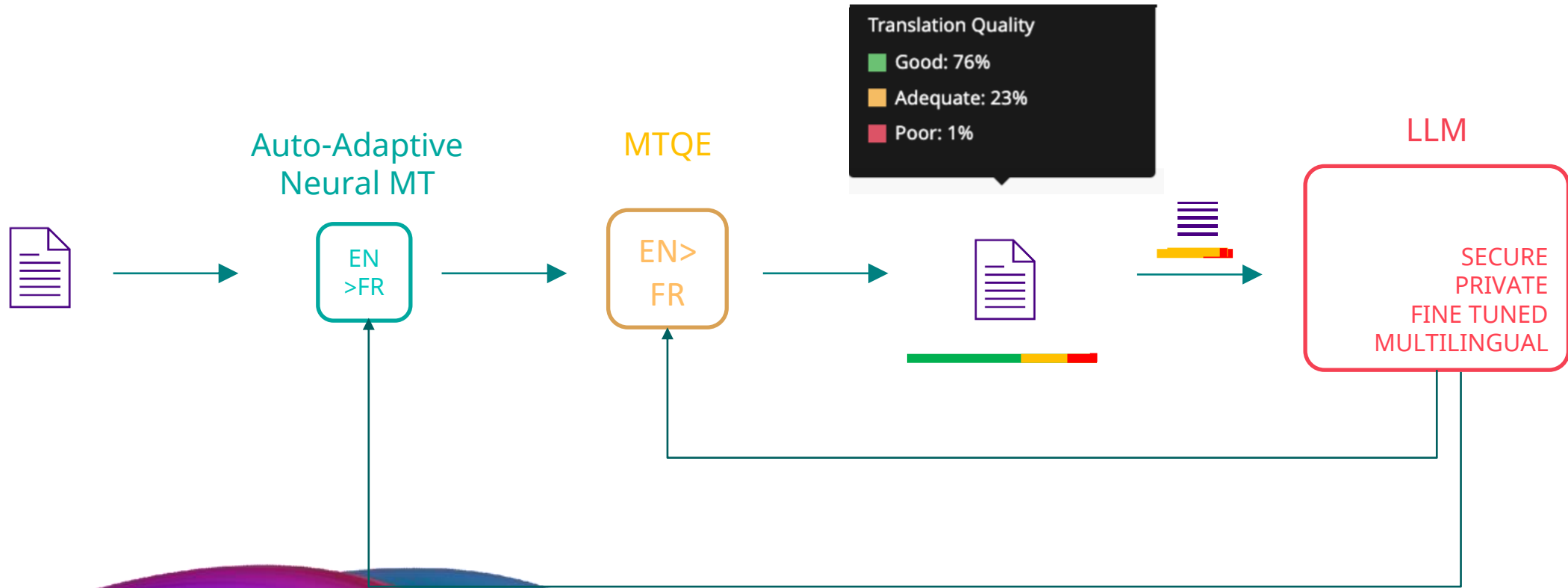
Technology and innovation combined with human interaction.

Language Weaver combines our AI technology with our linguistic expertise to help businesses solve the 'last mile' in their machine translation journey.



Language Weaver. The last mile in machine translation.

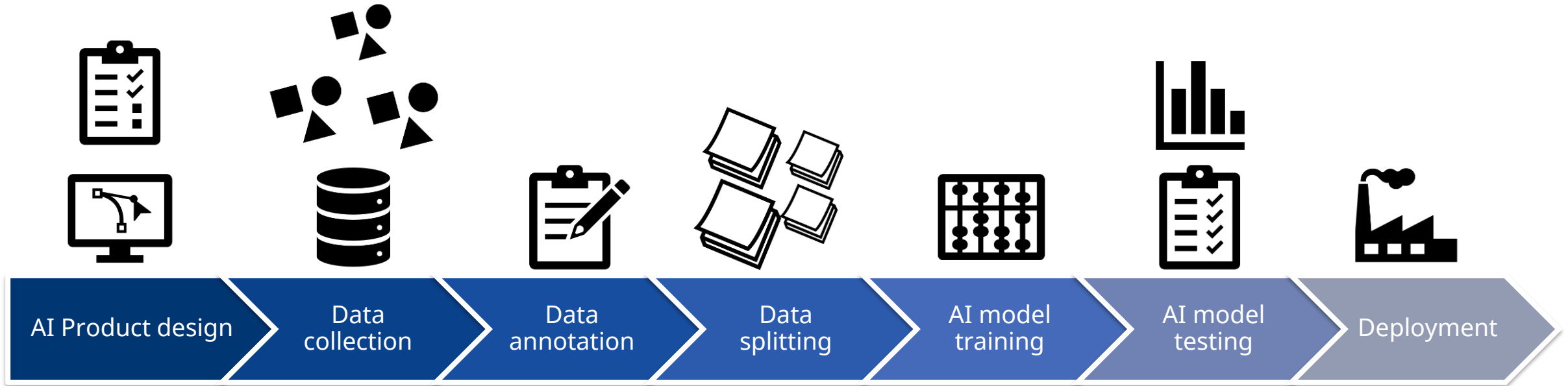
RWS Language Weaver





The AI product journey

A cost and data intensive process





It's all about the DATA!

What?

 High quality

 Source-verified

 Detailed metadata

 Verified compliance with regulations

 "One-stop-shop"?

Where?

Current efforts include:

- Uploading data to Huggingface
- Populating websites (research-driven)
- Linguistic infrastructure projects

How?



How was the data collected?



How can the data be used? / What license does it have?

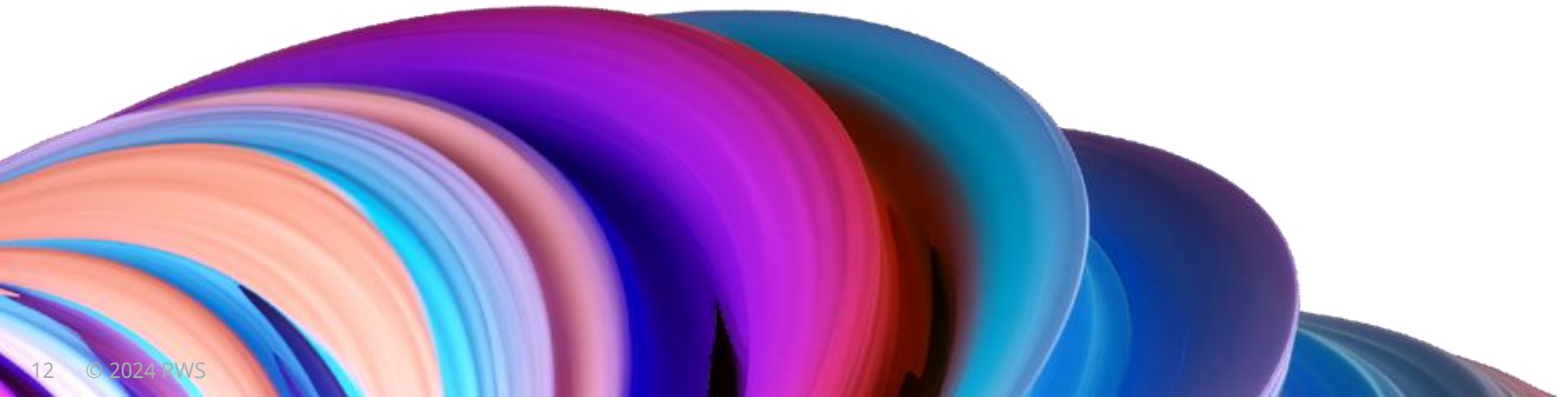


How can we ensure the right use?

Keeping track internally!

Companies need data repositories that:

- Are flexible to accommodate new tasks
- Allow to keep track of data
- Clearly separate training/test data
- Are populated with relevant metadata (licensing, use, etc.)





QUESTIONS?



THANK YOU



RWS Holdings plc is a unique, world-leading provider of technology-enabled language, content and intellectual property services. Through content transformation and multilingual data analysis, our combination of AI-enabled technology and human expertise helps our clients to grow by ensuring they are understood anywhere, in any language.

Our purpose is unlocking global understanding. By combining cultural understanding, client understanding and technical understanding, our services and technology assist our clients to acquire and retain customers, deliver engaging user experiences, maintain compliance and gain actionable insights into their data and content.

Over the past 20 years we've been evolving our own AI solutions as well as helping clients to explore, build and use multilingual AI applications. With 45+ AI-related patents and more than 100 peer-reviewed papers, we have the experience and expertise to support clients on their AI journey.

We work with over 80% of the world's top 100 brands, more than three-quarters of Fortune's 20 'Most Admired Companies' and almost all of the top pharmaceutical companies, investment banks, law firms and patent filers. Our client base spans Europe, Asia Pacific, Africa and North and South America. Our 65+ global locations across five continents service clients in the automotive, chemical, financial, legal, medical, pharmaceutical, technology and telecommunications sectors.

Founded in 1958, RWS is headquartered in the UK and publicly listed on AIM, the London Stock Exchange regulated market (RWS.L).

For further information, please visit: www.rws.com.

© 2024 All rights reserved. Information contained herein is deemed confidential and the proprietary information of RWS Group*.
*RWS Group shall mean RWS Holdings plc for and on behalf of its affiliates and subsidiaries.