

## Massively Multilingual Conference & Expo Program.

Tuesday - 11

San Jose, CA, USA | 11-12-13 October 2022

## Set up of Expo 15:00 - 18:00

18:00

**Opening Reception** 

at Gateway Ballroom in Hilton San Jose Hotel, sponsored by SYSTRAN

## **Registration Opens** 08:00 - 09:00 Coffee and tea (expo is open)

Wednesday - 12

Massively Multilingual Models and the state of Machine Translation in 2022 What everybody should know about encoders, decoders, transformer models and other key concepts in the world of MT before the discussion starts. A mini-lecture by John DeNero.

The goal of the Microsoft XYZ initiative is to train joint representations across multiple

09:40

**Everything All At Once: Machine Translation at the Convergence** by Hany Hassan Awadalla

help us create a more powerful human-centric AI that can learn to speak, hear, see, and understand the way humans do. The goal is to have pre-trained models that can jointly learn representations to support a broad range of downstream AI tasks. Are we there? Not yet. The ZCode project combines multi-lingual and multi-task training to support a wide variety of NLP

languages, multiple modalities and multiple tasks, covering NLP, Translation, Speech and Vision. At Microsoft, we believe cross-domain transfer learning spanning languages and modalities will

training along the other axes of our XYZ vision. Massively Multilingual Panel Discussion with Hany Hassan Awadalla (Microsoft), Paco Guzmán (Meta), Macduff Hughes (Google), moderated by John DeNero (Lilt) and Jaap van der Meer

What was it again? MT is a simple sum of models and data. During the pandemic over the last

without language barriers? Or will we have to fall back on the finesse of good quality data work

three years, the tech world has invested massively in magic models. Following the keynote opening address - Everything All At Once - the leaders of the four big tech MT platforms will make up the balance. Will the brute force of the massive models do it all and bring us a world

to tune the models and close the quality and the credibility gap? Each of the speakers will share a perspective on the prediction of massively multilingual MT, then opening up for a panel discussion with John DeNero and Jaap van der Meer and Q&A with the audience.

(expo is open)

**Refreshment Break** 

11:30

11:35

12:30 **O** 

15:00 - 15:30

11:00 **O** 

11:30

Language accessibility programs made easy and cost-efficient, Konstantin Savenkov (Intento)

Introduction, by Anne-Maj van der Meer (TAUS)

World-Readiness Contest (Part I)

11:41 Quality prediction for hybrid translation, Adam Bittlingmayer (ModelFront) Human or machine, quality or scale? Quality prediction is the key to hybrid translation workflows - the only way to provide quality AND scale. Now it's available as an API that supports more than 10,000 language pairs.

inclusive design choices to anticipate future market readiness, d) non-intrusive measurability of success. On Wednesday afternoon, all attendees will vote for their favorite and the winner will be

Language accessibility plays a critical role in global enterprises. It ensures everyone has access to all relevant

The Church of Jesus Christ of Latter-day Saints has a concrete mandate to convey its message to the very last speaker on Earth. While we currently operate in about 150 languages, it would take us centuries to expand into all other languages at our present rate. This presentation covers our approach for tapping into existing talent to

**DeMT™ Evaluation Report** 12:00

Anderson Vaz (TAUS) and Achim Ruopp (Polyglot Technology)

presentation will discuss TAUS Data-enhanced Machine Translation (DeMT™), the essential layer of curated high-quality and domain-specific data that TAUS puts on top of the MT engines in order to deliver near-human quality output. The speakers will present the findings of the TAUS DeMT™ Evaluation Report, which covers an in-depth analysis of the training performed on various prominent MT engines.

On Evaluating MT, human parity and the limits of MT 14:00 with Markus Freitag (Google), Gretchen Markiewicz (Raytheon BBN) and Philipp Koehn (Facebook AI), moderated by Olga Beregovaya (Smartling) After discussing the training models and the promises of the data work helping us to reach massively multilingual engines, it's time to talk about evaluation. Without measurement, it's very

The translation industry at large has not yet come fully to terms with the full power of MT. Most players look at MT as a plug-in or an add-on to their existing translation processes. That's not how we see the future of translation at TAUS. We advocate fundamental redesigns of translation workflows with MT at the center. In this part I of The Business Perspective we focus on the explosion of new use cases for multilingual communications that were unthinkable until only recently. **TBC** 15:00 O

> the only obstacle to true universal communication. In the past, people either needed to find a common language to communicate in, often in a sub-optimal manner, or needed the help of human translators and interpreters. If they did not have any language in common and human translators were either not available or not affordable to the parties involved, communication was impossible. Speech translation technology now offers a solution for overcoming this language barrier in unified communication solutions. For this, it has to be tailored to different scenarios, such as low-latency online translation or off-line translation of recordings; it has to be robust to different accents, channels and microphone conditions; and it has to cover as many languages as possible, and, in the case of translation, as many language pairs as possible, out of the 7,000 existing languages in the world. In this talk we will present how Zoom is introducing

Improving meeting productivity and efficiency with

**Panel Conversation** with Watson Srivathsan (Amazon), Sebastian Stüker (Zoom) and Murali Nathan (Avery Dennison). Moderated by Anne-Maj van der Meer and Antoine Moreau

machine and speech translation technology into its communication products to overcome the

Automating cross-lingual communications at Avery Dennison

Thursday - 13

Coffee & Tea

(expo is open)

Welcome and reflect on Day I 09:00 by Jaap van der Meer

The Business Perspective (Part II)

**Al-enabled Translation Workflows** 09:30 with Zhen Chao (VMware), Wayne Bourland (Dell), Loic Dufresne de Virel (Intel), moderated by Renato Beninatto and Anne-Maj van der Meer

Quantifying Quality – Translation Memories, Mei Zheng (Smartling)

scores generated by our approach are consistent with our anecdotal findings.

**AI-Enabled Translation** 

Moreover, services change constantly, as MT providers work hard to improve their offerings. What is needed is a way to independently assess quality, monitor quality and determine if quality meets your needs. With MT Decider Polyglot Technology offers a comprehensive set of evaluation reports, real-time MT monitoring and evaluation services - for generic content, domain-specific content (as TAUS DeMT Evaluate in collaboration with TAUS) and, most importantly, your content. Vendor-independent evaluation allows you to confidently choose the MT solution that is right for you.

You're not going to reach the last person in the world on your own – you'll need tools. In that toolbox, you most

translation memory largely depends on its quality. Here we present our two-parts approach to quantify the quality of translation memories so that you're left with only the most accurate translations. Our approach consists of a

quantitative evaluation of TMs' relevance in addition to a string-by-string level quality analysis. We found that the

No Language Left Behind: Scaling Human-Centered Machine Translation

Most of today's machine translation (MT) models work for mid- to high-resource languages—

research today. However, such efforts have coalesced around a small subset of languages,

contextualizing the need for low-resource language translation support through exploratory

language barrier while ensuring safe, high quality results, all while keeping ethical

considerations in mind? In No Language Left Behind, we took on this challenge by first

leaving most low-resource languages behind. Driven by the goal of eradicating language barriers on a global scale, machine translation has solidified itself as a key focus of artificial intelligence

leaving behind the vast majority of low-resource languages. What does it take to break the 200-

certainly should have a translation memory (TM)! However, the degree of benefit that one can reap from a

for many language pairs varies widely, particularly when scaling to dozens of languages. Mind the gap!

**Beyond the Traditional Languages** 11:30 - 12:00

11:30 O

11:40

12:00 **O** 

14:00

14:05

14:11

16:00

16:50

interviews with native speakers. Then, we created datasets and models aimed at narrowing the performance gap between low and high-resource languages. We propose multiple architectural

**Human Language Project** 

Simona Beccaletto (TAUS)

World-Readiness Contest (Part III)

A practical way to expand software locales, Todd Flaska (Lingoport)

Behind project.

Vedanuj Goswami (Meta)

**Building Out Living Lexicons** 11:50 Casper Grathwohl (Oxford Languages) Casper will provide an overview of the Oxford Languages initiative to build out living lexicons in 40 low-resource languages. Opportunities for language expansion 12:00 Panel discussion with Simona Beccaletto, Vedanuj Goswami, and Casper Grathwohl (Oxford Languages). Moderated by Solen Aslan and Grainne Maycock **Lunch Break** 

for developers) followed by an "in context" review. Plus, the reviewer can be traditional LSP/translators, field personnel, resellers/distributors, or even customers. This eliminates the issue of sourcing translators in far flung languages or dialects. So now there is no excuse: "I expect your software to work in my language". Lexicala's expert multi-layer, cross-lingual data resources for massive, refined machine translation, Karni Bernad

Lexicala develops multi-layer cross-lingual data networks by converging manually created content with smart automated processes topped by human curation. Our model applies one overall framework and its underlying technical infrastructure to generate expert bilingual and multilingual parallel corpora for refined machine

Cohen (Lexicala)

(expo is open)

**NLP Panel** 14:30 with András Aponyí (TAUS), Adam Bittlingmayer (ModelFront), Sunil Mallya (Flip AI). Moderated by JP Barraza

and trends in Natural Language Processing.

Moderated by Paul Mangell (Alpha CRC)

**Any Questions?** 

**Refreshment Break** 15:30 **O** (expo is open)

This panel of highly-experienced NLP engineers will be diving deep into the latest innovations

of topics around the localization industry, in an informal and personalized format which allows any attendee to ask questions, and the panel of experts to respond with their own take on the subject. So whether it's around AI, or NMT, or process and workflow management, or resourcing and global supply chains, this session explores the matter with unscripted, personal responses.

Any Questions? (based on the BBC current affairs program) is an opportunity to explore a range

Welcome & Program Overview 09:00 by Jaap van der Meer **Keynote Session** 09:10 - 10:10 Introduction to the session and presenting the speakers 09:10 by Jaap van der Meer 09:20

> and NLG tasks, including translation. I will discuss the challenges of shipping these very large models into a production cloud API as well as touch upon recent progress towards multi-model 10:10

For the first time ever, TAUS organizes the World-Readiness contest. In this contest, we'll be hearing from companies about their strategies and approaches to reach the very last human being on the planet in their own native language. Each presenter presents a short pitch which will be evaluated by the attendees over the course of two days. The criteria include a) the extent and specifics of new human language/domain coverage, b) interoperability of tech solutions to ensure broad take-up, c)

announced.

accelerate our pace.

**Lunch Break** 

(expo is open)

knowledge and can communicate with peers without language barriers. Can the traditional approach to translation management bridge this gap with a quick turnaround and tight budget? Intento team will show examples of strategies successfully implemented by global companies. Unto All the World, Zak Nyberg and Ana De Agostini (The Church of Jesus Christ of Latter-Day Saints) 11:47

MT customization essentially requires two elements: an MT model and training data. This

difficult to make any advancements. So how do we know the models and the data enhancements are good enough? Evaluating MT is the holy grail. The speakers in this panel will be discussing evaluation methods and systems, the goal of reaching human parity and what the limits of MT are.

The Business Perspective (Part I)

Watson Srivathsan (Amazon)

language barrier for its users.

with Murali Nathan (Avery Dennison)

multi-language translation

with Sebastian Stüker (Zoom) Modern communication technology, such as offered by Zoom, has broken down barriers and is constantly connecting people from all parts of the world. The language barrier now remains as

15:10

15:20

16:00

16:15

17:00

08:30 - 09:00

09:15 - 09:30

10:30 **O** 

11:00

11:05

11:11

11:17

15:30 **Adjourn** Shuttle leaves for Regale Winery & Vineyard Dinner at Regale Winery & Vineyard valid dinner ticket needed. Purchase it here: https://buy.stripe.com/dR6g1fd8sdiu8Ug8ww

Renato Beninatto (Nimdzi Insights) Advances in AI provide new opportunities to reduce costs and improve the quality and availability of translation services. To automate and augment translation processes, the use of Al has become inevitable.

Report.

09:15

**Refreshment Break** 

(expo is open)

World-Readiness Contest (Part II) Mind the Gap, with Achim Ruopp (Polyglot Technology) With the widespread implementation of neural machine translation online machine translation services are often

considered interchangeable commodities. But in reality this isn't really the case: the translation quality of services

challenges and more insights about their digital transformations.

In this panel conversation, we'll be talking with enterprises who are in various stages of adopting

Al and ML in their traditional translation workflows. We will be hearing about the steps, the

In this Part II of the Business Perspective, we are talking with heads of localization and globalization who face

the challenge of adapting to the new reality of Al-Enabled Translation as painted by Gartner in their Market

SYSTRAN Meeting Translator - Real-time STT with HILT review, JP Barraza (Systran) SYSTRAN Meeting Translator is a real-time speech-to-text captioning solution for secure meetings/events with inline human-in-the-loop edit/review capabilities.

and training improvements to counteract overfitting while training on thousands of languagepairs/tasks. We evaluated the performance of over 40,000 different translation directions using a human-translated benchmark, Flores-200, and combined human evaluation with a novel toxicity benchmark covering all languages in Flores-200 to assess translation safety. Our model achieves state-of-the-art performance, laying important groundwork towards realizing a universal translation system. We open source all contributions from the No Language Left

incorporate locales that are outside the mainstream. Their concerns are both about cost and quality for nonmainstream locales. Lingoport solves this through a unique combination of immediate machine translation (great

For many, the software User Interface (UI) is the face of the company to their customers so they are reluctant to

translation for any language – as source, target, or pivot – relying on deep lexical mapping and including sense alignment, domain classification, morphology and annotation services.

Wrap-up & Announcement of World-Readiness Winner More information about the Massively Multilingual Conference & Expo 2022 can be found here: https://www.taus.net/events/conferences/taus-massively-multilingual-conference-expo-2022/