LIONBRIDGE

Delivering the Next Phase of Localization Value

SPEAKER:

VINCENT HENDERSON
VP Strategy, Language AI
Lionbridge

AI must be used responsibly

TRANSPARENT RELIABLE USEFUL SCALABLE TIMELY

TRUST

Achieve goals | Achieve more

A framework to ensure that AI actually helps us **achieve our goals**

ROI ENGAGEMENT AUDIENCE CONTROL HUMAN-IN-THE-LOOP

REACH

Define content goals | Achieve content goals

A framework to ensure that we know **what our goals are** and we control them

AI must be used responsibly

TRUST

TRANSPARENT RELIABLE USEFUL SCALABLE TIMELY

Achieve goals | Achieve more

TRANSPARENT

AI systems' operations and **decision-making** processes are **auditable** and **explainable**.

RELIABLE

AI systems **perform consistently**, are error and manipulation-resilient, and uphold high **safety standards**

USEFUL

AI systems help meet **specified objectives**, **adding value** to human endeavors.

REACH

ROI ENGAGEMENT AUDIENCE CONTROL HUMAN-IN-THE-LOOP

Define content goals | Achieve content goals

CONTROL

Automated gathering of engagement and content **quality data** drives decision-making on content specifications.

HUMAN-IN-THE-LOOP

The **AI-first** process is controlled, curated and **informed by human oversight** and insights to optimize every step.

AI must be used responsibly

TRANSPARENT RELIABLE USEFUL SCALABLE TIMELY

TRUST

Achieve goals

Achieve more

SCALABLE

AI systems can handle **increasing workloads** and **maintain performance** and quality.

TIMELY

AI systems are responsive, updated, and provide insights and work product in a timely, relevant manner.

ROI ENGAGEMENT AUDIENCE CONTROL HUMAN-IN-THE-LOOP

REACH

Define content goals

Achieve content goals

ROI

The investment in your global content matches expected returns and delivers business outcomes.

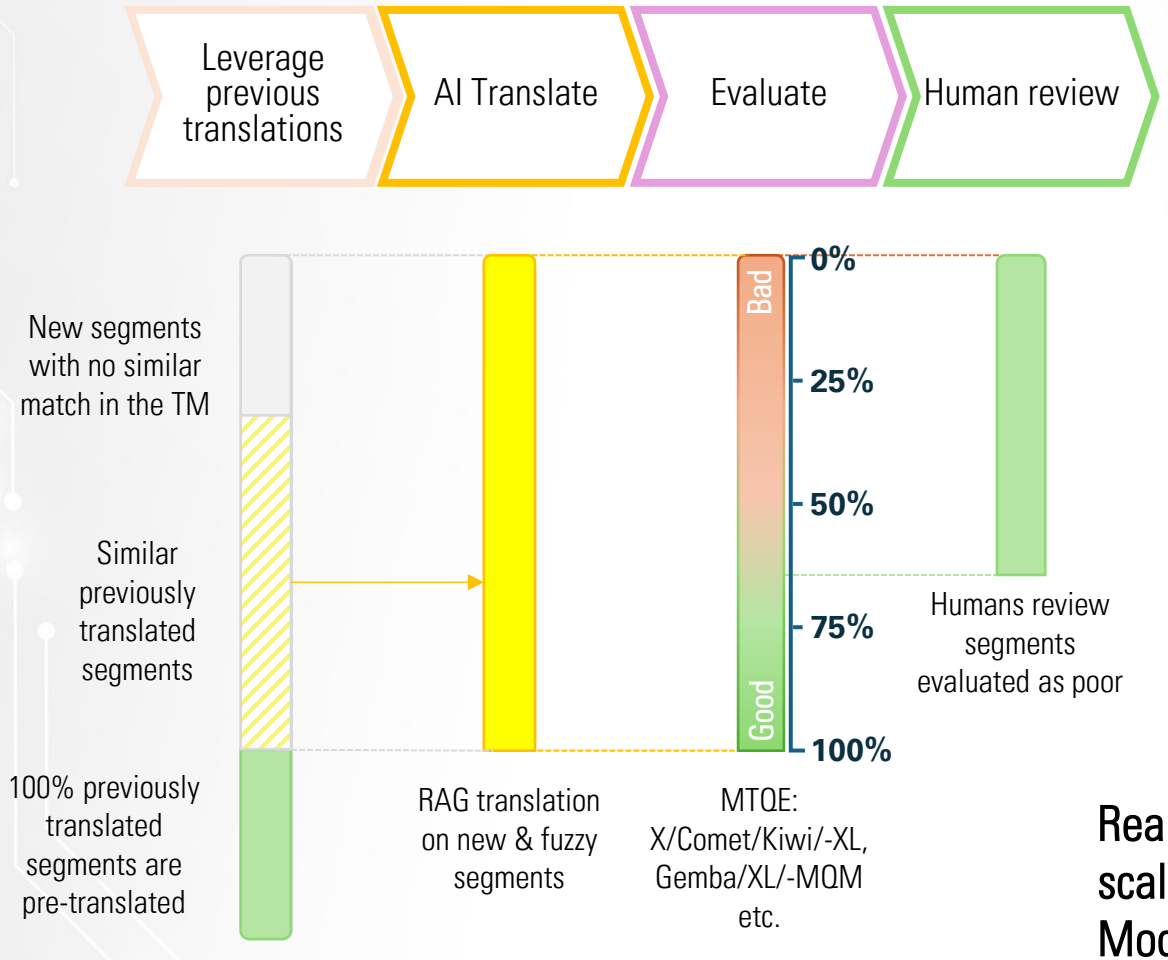
ENGAGEMENT

The content is specifically adapted to engage your global targets on specific goals.

AUDIENCE

The content resonates with each of your audiences in global markets on a cultural, community and demographic level.

THE CANONICAL SOTA AI TRANSLATION WORKFLOW: RAG + MTQE



Various MTQE approaches have various advantages and drawbacks: speed, cost, n-shot examples, accuracy.

In production workflows, we can only use reference-free frameworks. The average published **accuracy** of these generic reference-free models is ~85%.

These models' published accuracy evaluations only deal with untagged, unformatted text and ignore **real-world translation** quality assets (glossaries, style guides, quality targets, tagging schemes).

RAG translation solves some of these issues by boosting translation compliance at the source, but only to some extent (~30% MT quality boost).

Real-world, professional, industrial translation at scale requires more engineered approaches. Modern LLMs give us cognitive features that we can engineer in our software automation.

DETAILED BREAKDOWN OF AN ENGINEERED AI-FIRST TRANSLATION FLOW

Establish quality expectation | Leverage previous translations | AI Translate | AI enhance & evaluate | Human review

New segments with no similar match in the TM

Similar previously translated segments are identified for reference

Q1
Q2
Q3
Q4
Q5

Target accuracy

100% previously translated segments are pre-translated

We start from an expected quality goal

Initial RAG translation on new segments

for review

locked

AI post-edit and evaluate segments for review. 100% matches on low-confidence TMs can be included.

Quick check | Light review | Light edit | Full edit

effort

accuracy

Accuracy of AI evaluation (known)

Human editing achieves different defect removal accuracy with level of effort

The maximum level of editing effort applied to segments "for review" to reduce a high rate of defects

If the "locked" accuracy is below the target accuracy, a moderate level of editing effort for "locked" segments is used to cleanup a small number of defects

Target accuracy

The algorithm selects the edit level required to reach the accuracy target

Humans review segments identified by AI

Final segments ready for publishing

QUALITY ASSURANCE AND ONGOING BENCHMARKING

Ramp-up

Production

Continuous improvement

Process, improve & customize

Leverage previous translations

AI Translate

AI enhance & evaluate

Human review

Sampling & validation

Update reference assets

Process jobs for a period of time with 100% review to gather initial KPI data and increase AI performance

Similar previously translated segments are identified for reference

100% previously translated segments are pre-translated

Initial RAG translation on new segments

Update the APE accuracy benchmarks

for review

locked

AI post-edit and evaluate segments for review

Humans review some segments

Published segments

Ongoing sampled quality assurance and validation of new content and legacy TM segments

Add / update validated content to reference assets based on quality rules

CONTINUOUS IMPROVEMENT & COST REDUCTION

Ramp-up

Production

Continuous improvement

Process, improve & customize

Leverage previous translations

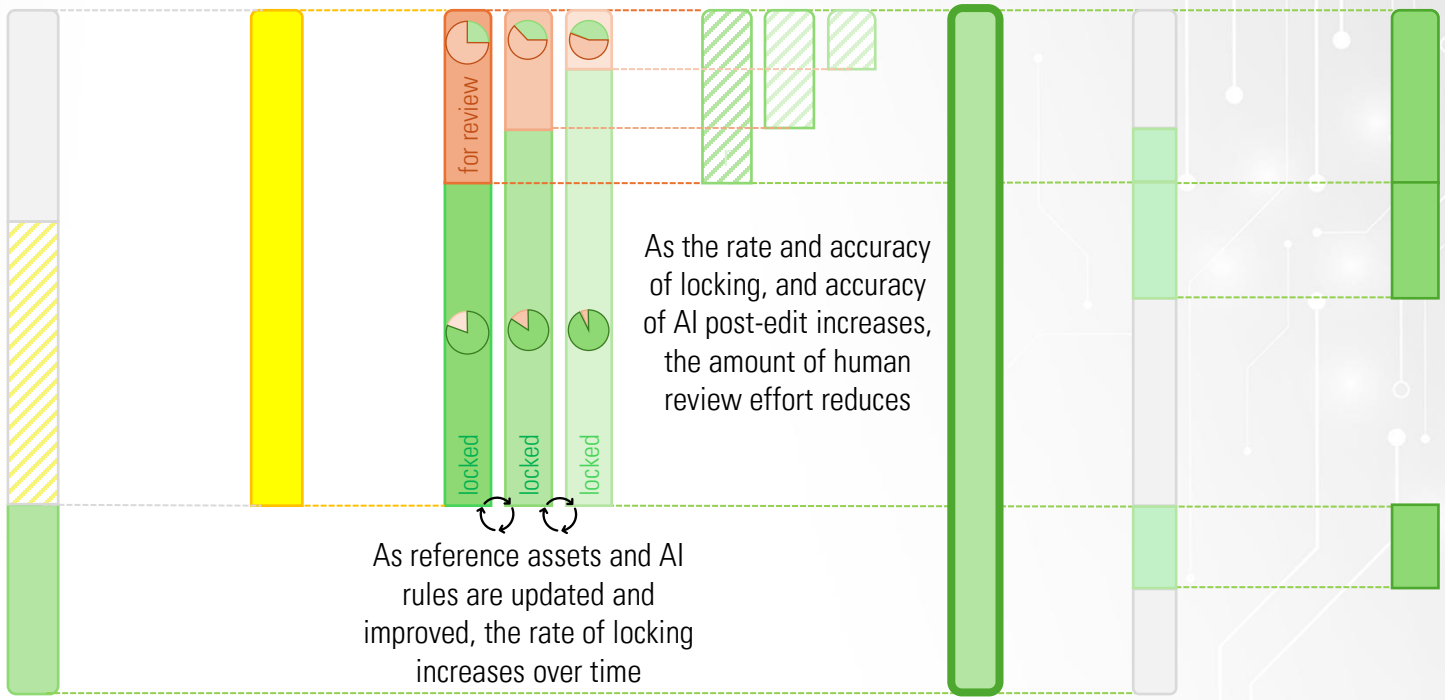
AI Translate

AI enhance & evaluate

Human review

Sampling & validation

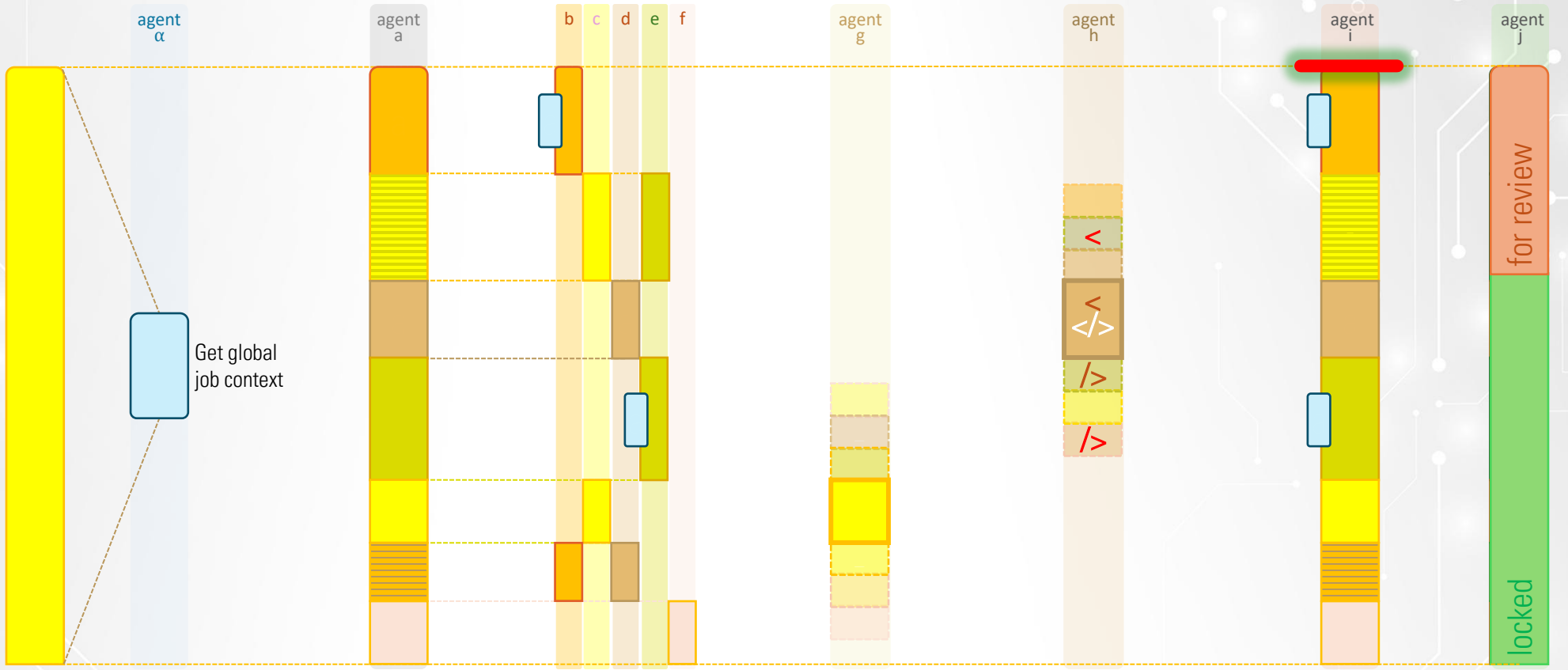
Update reference assets



ENGINEERED AGENTIC TRANSLATION ENHANCEMENT

AI Translate

AI enhance & evaluate



RAG-summarize job content and analyze for key non-glossary terminology

Categorize trans-units for specific issues: numbers, units, acronyms, lists, tags, DNTs, headings, glossary, etc…

Provide ONLY the relevant instructions, examples and context with each trans-unit

Resolve anaphoric and local context issues: pronouns, gender agreement between references, synonyms, etc.

Apply semantic and structural tagging correctly in target segment to reflect source tagging

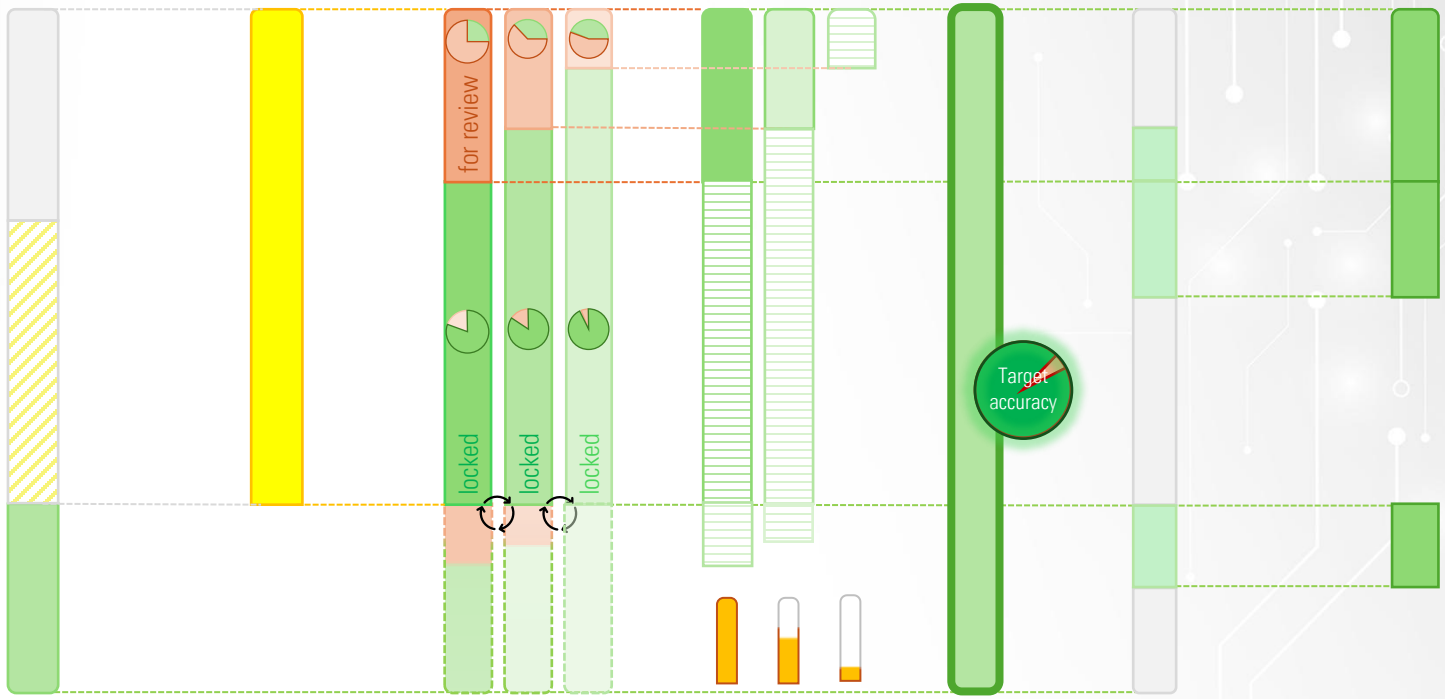
Evaluate correctness on critical localization issues

Annotate trans-units for review

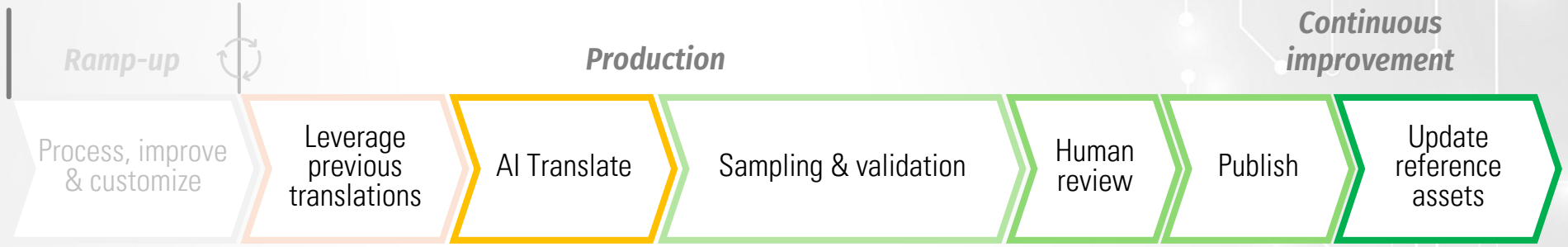
CONTINUOUS IMPROVEMENT & COST REDUCTION

Reducing the cost of good quality translation is no longer a question of having "better MT engines" or "better LLMs" or any one-size-fits-all "better QA". We have reached the state where **the language competence of the tech is no longer the issue**.

The key to good quality translation has always been about **whether translation specifications are well defined** and applied consistently (error definitions, content audience and engagement goals, what style rules apply, etc.).



USING AUTO-LQA TO VALIDATE TRANSLATIONS

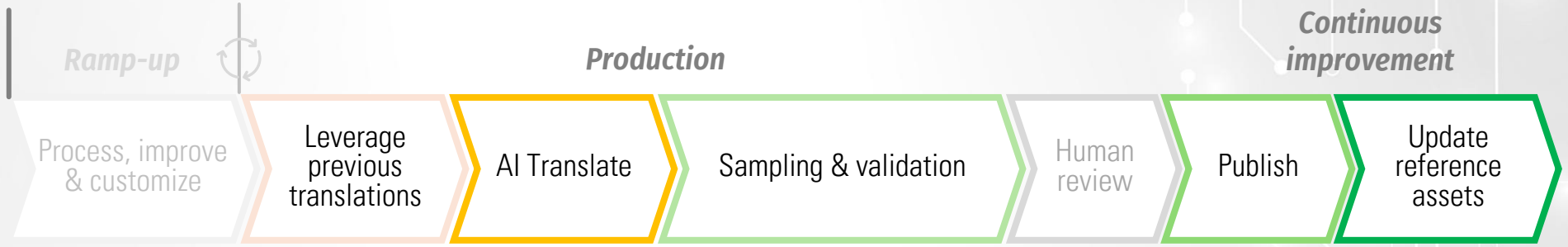


AutoLQA can take any bilingual segmented XLIFF corpus as its source.

If another translation process outputs a corpus that is a candidate for publication where there is no expectation of a consistent need for human editing, AutoLQA can be used to catch the **occasional** failed job for review before publishing.

It could be particularly relevant for on-the-fly online publishing where speed is of the essence and where corrections can be easily made after initial publishing.

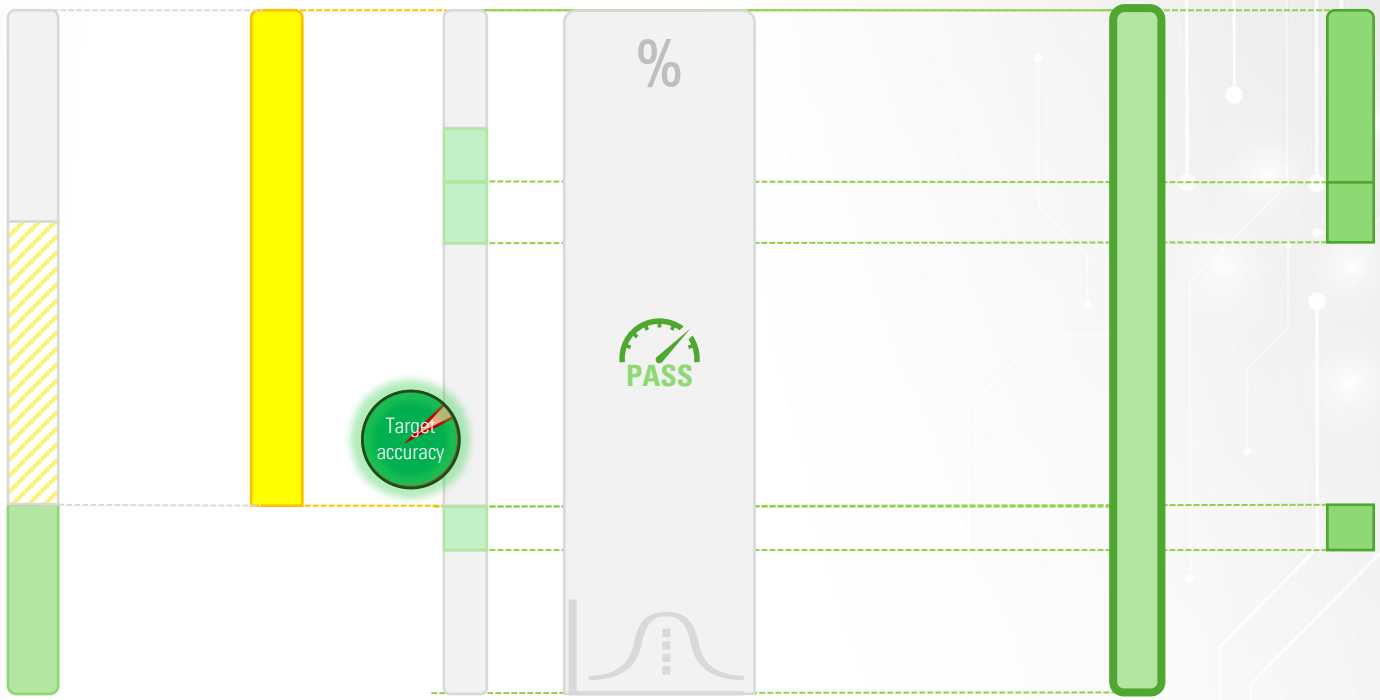
USING AUTO-LQA TO VALIDATE TRANSLATIONS



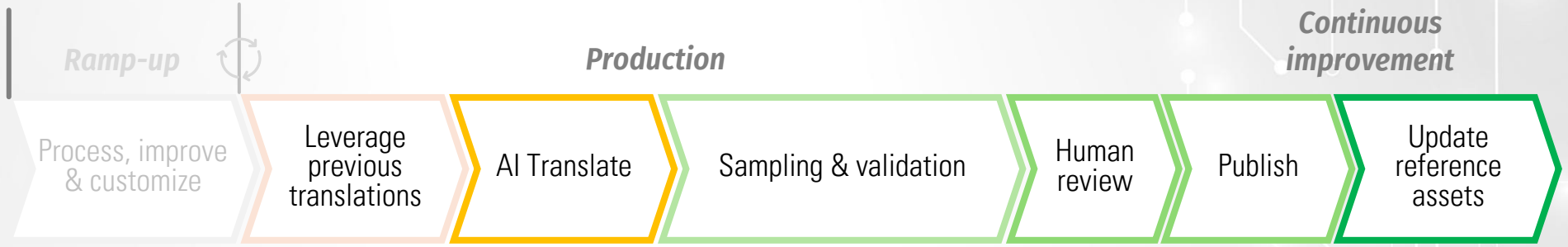
AutoLQA can take any bilingual segmented XLIFF corpus as its source.

If another translation process outputs a corpus that is a candidate for publication where there is no expectation of a consistent need for human editing, AutoLQA can be used to catch the **occasional** failed job for review before publishing.

It could be particularly relevant for on-the-fly online publishing where speed is of the essence and where corrections can be easily made after initial publishing.



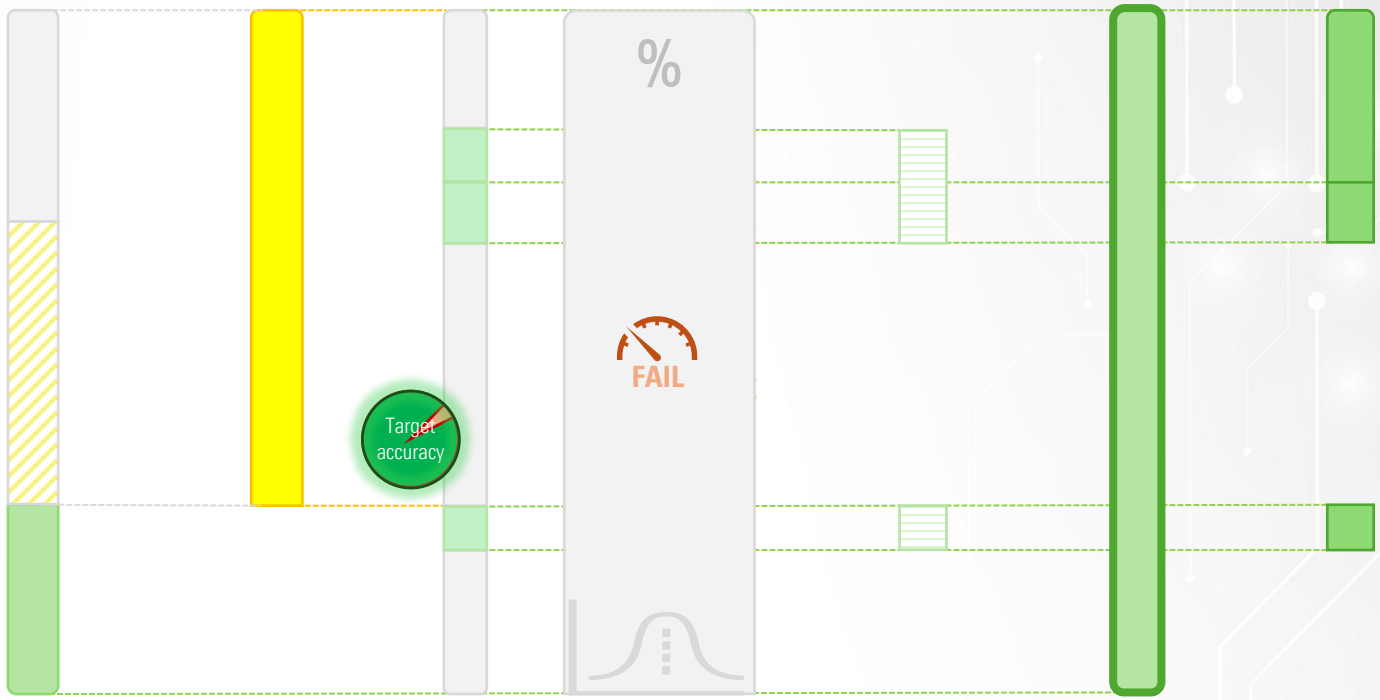
USING AUTO-LQA TO VALIDATE TRANSLATIONS



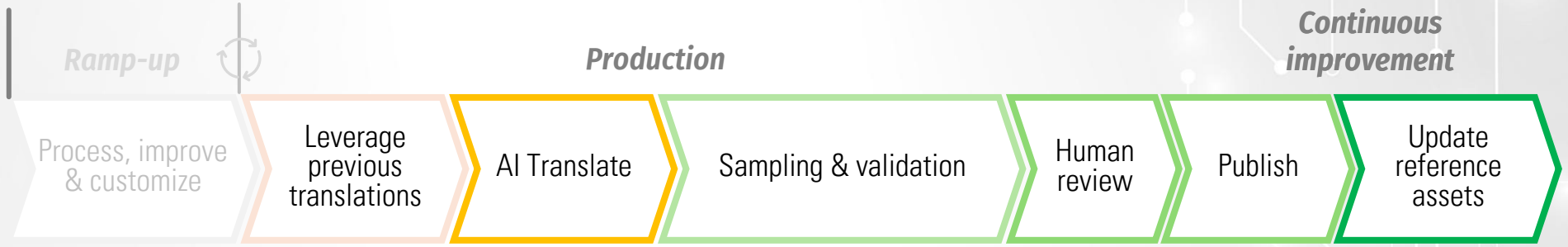
AutoLQA can take any bilingual segmented XLIFF corpus as its source.

If another translation process outputs a corpus that is a candidate for publication where there is no expectation of a consistent need for human editing, AutoLQA can be used to catch the **occasional** failed job for review before publishing.

It could be particularly relevant for on-the-fly online publishing where speed is of the essence and where corrections can be easily made after initial publishing.



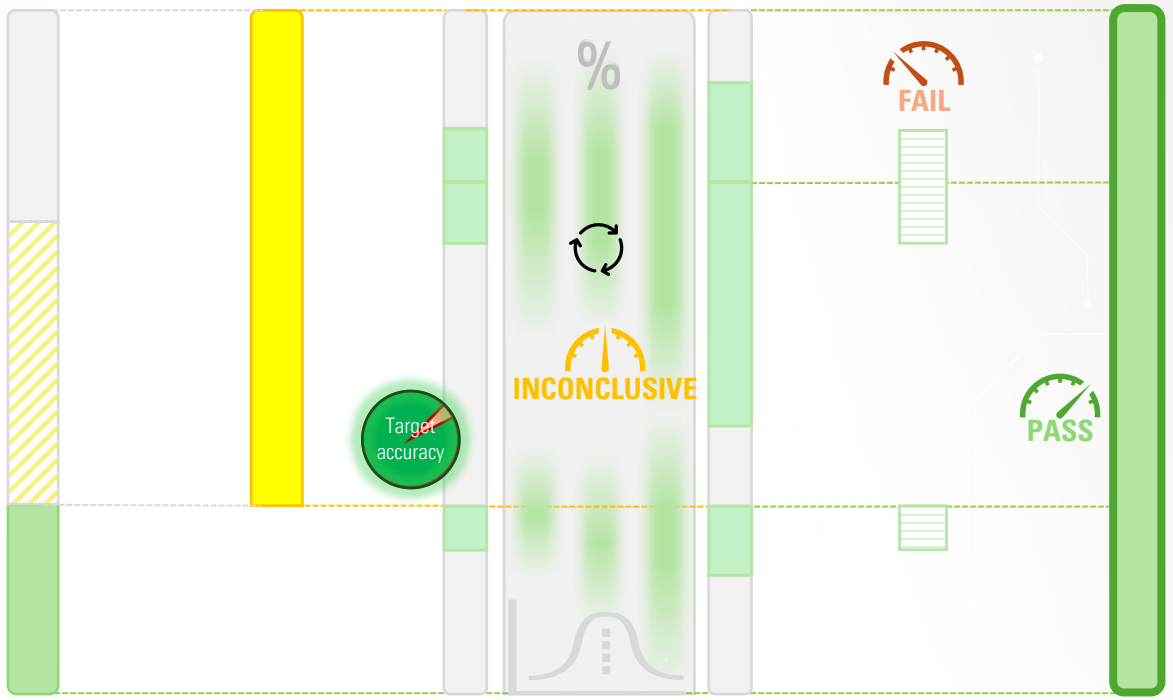
USING AUTO-LQA TO VALIDATE TRANSLATIONS



AutoLQA can take any bilingual segmented XLIFF corpus as its source.

If another translation process outputs a corpus that is a candidate for publication where there is no expectation of a consistent need for human editing, AutoLQA can be used to catch the **occasional** failed job for review before publishing.

It could be particularly relevant for on-the-fly online publishing where speed is of the essence and where corrections can be easily made after initial publishing.



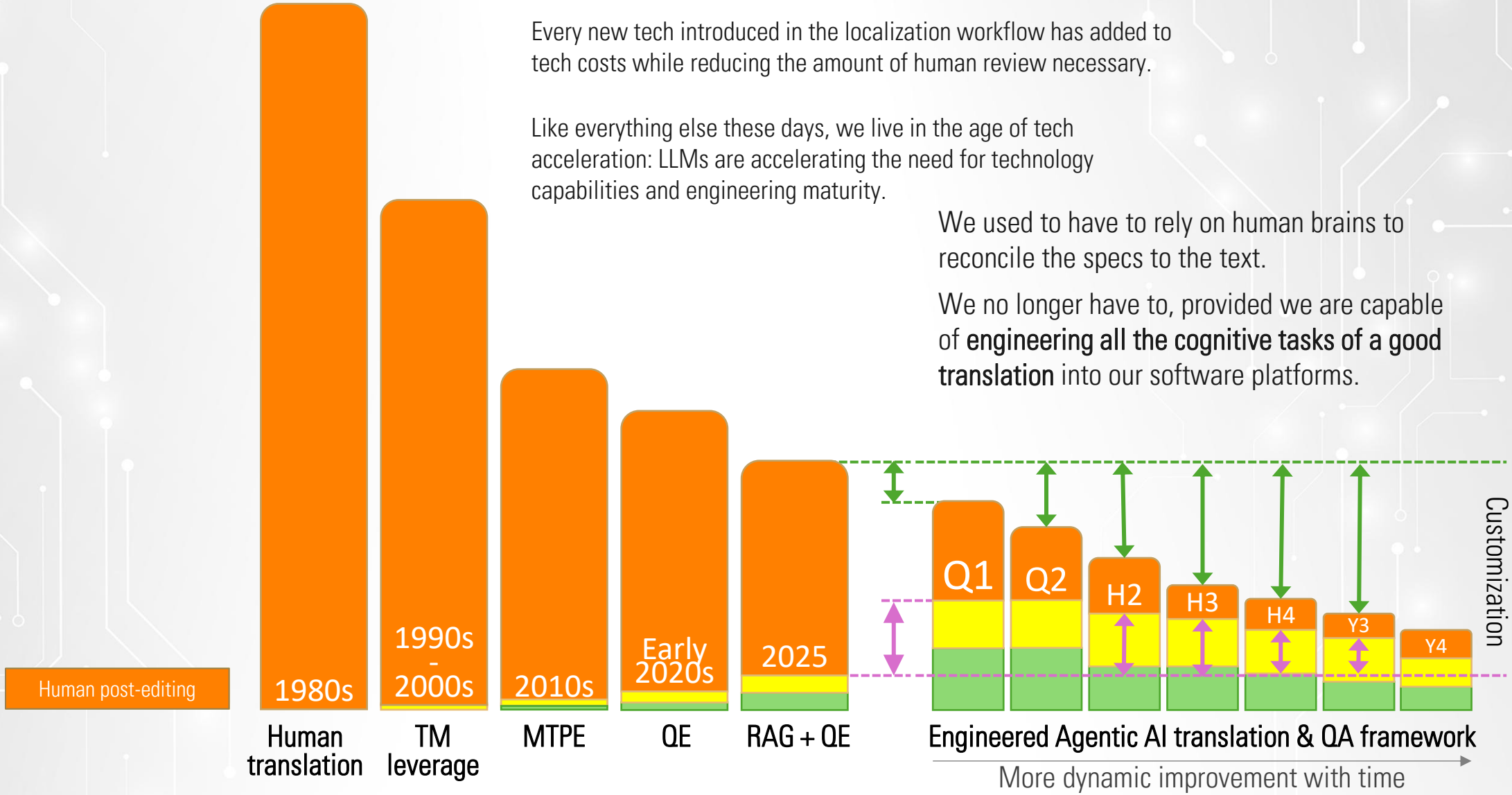
ENGINEERING WITH LLMs FOR CONTINUOUS IMPROVEMENT & COST REDUCTION

Every new tech introduced in the localization workflow has added to tech costs while reducing the amount of human review necessary.

Like everything else these days, we live in the age of tech acceleration: LLMs are accelerating the need for technology capabilities and engineering maturity.

We used to have to rely on human brains to reconcile the specs to the text.

We no longer have to, provided we are capable of **engineering all the cognitive tasks of a good translation** into our software platforms.



LIONBRIDGE

LIONBRIDGE
THANK YOU
We break barriers and build bridges to new opportunities. Everywhere.