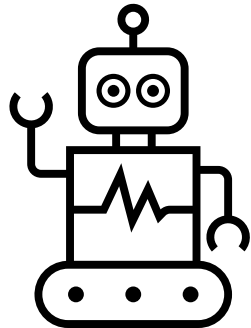
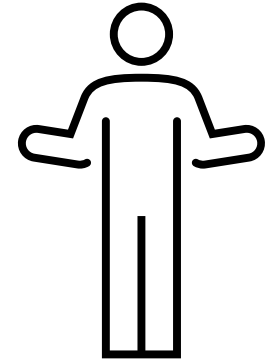


# 機械翻訳は どこまでも 可能か？

2024年12月3日(火) 10:10~10:50 (40分/質疑応答含む) (国研) 情報通信研究機構 隅田英一郎

「機械翻訳はどこまで可能か」について教えてください。

「機械翻訳はどこまで可能か」というタイトルの本について教えてください。



「機械翻訳はどこまで可能か」は、長尾真（ながおまこと）著の本で、1986年に岩波書店から出版されました<sup>1</sup><sup>2</sup>。この本は、機械翻訳の可能性と限界について詳しく解説しています。

#### 主な内容

- **機械翻訳の歴史:** 機械翻訳の発展の歴史とその背景について説明しています。
- **技術的な側面:** 機械翻訳システムの技術的な詳細や、どのようにして翻訳が行われるかについて解説しています。
- **課題と限界:** 機械翻訳が直面する課題や限界についても触れています。
- **未来の展望:** 将来的に機械翻訳がどのように進化する可能性があるかについての考察も含まれています。



この本は、機械翻訳に関心がある方や、技術の進歩に興味がある方にとって非常に有益な内容となっています<sup>1</sup><sup>2</sup><sup>3</sup>。

他に知りたいことがあれば教えてください。

詳細情報 [1 amazon.co.jp](#) [2 books.rakuten.co.jp](#) [3 jstage.jst.go.jp](#) [4 e-hon.ne.jp](#) +1 その他

👍 🗨️ 📄 ⬇️ 🔊 1 / 30 応答 AIによって生成されたコンテンツは誤っている可能性があります

**(1986年) 温故知新～課題とコーパスベース技術～**

**(2021年ごろまで) ニューラルネットによる到達点**

- 文書翻訳の自動化
- 通訳の自動化
- 語学への影響

**(1986年) 温故知新2～課題と生成AI～**

**(2022年～2025年以降) 翻訳/通訳はステージが上がる**

- コンテキストの活用
- 超多言語化
- 機械翻訳の進歩は後戻りしない

溫故

知新

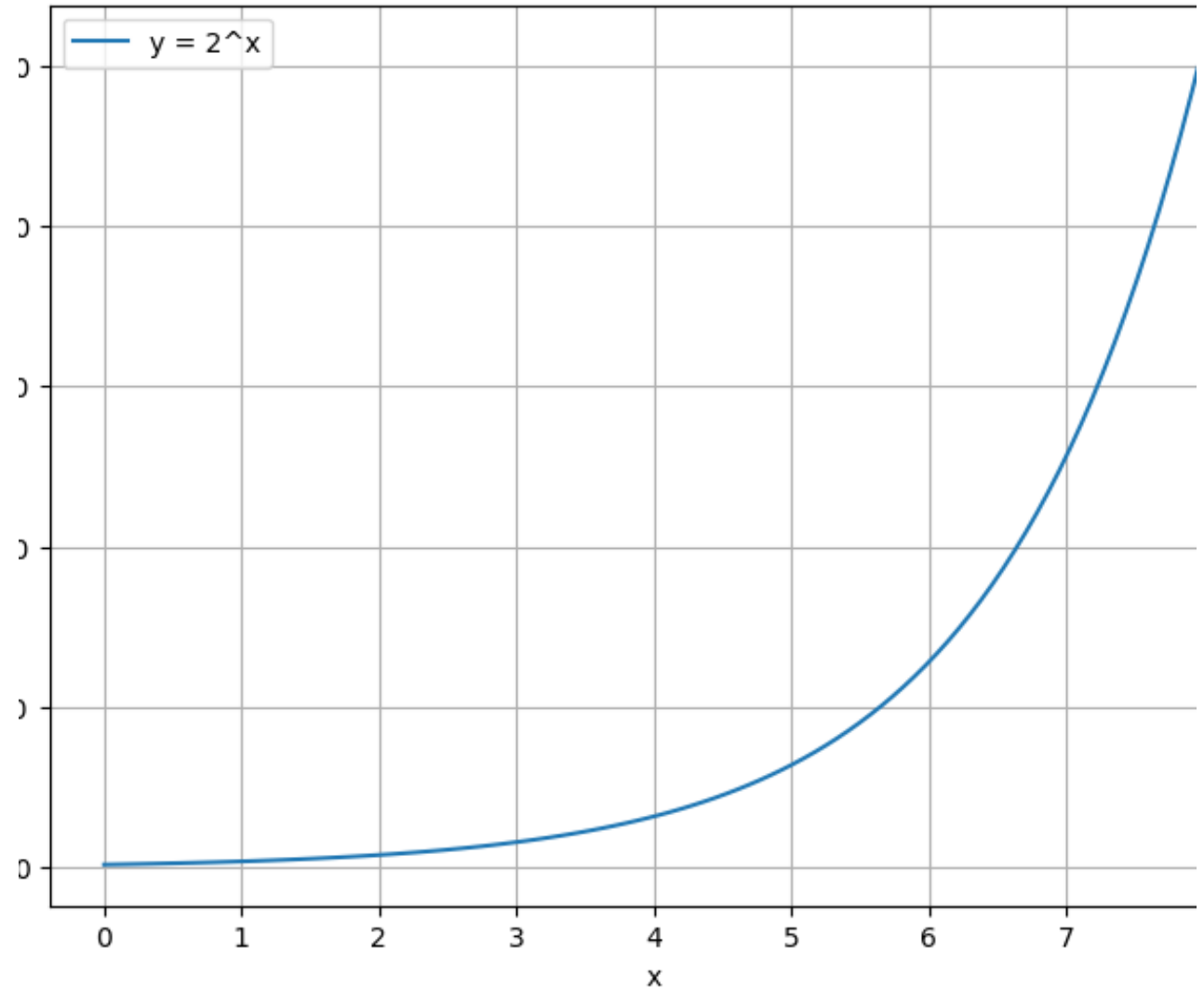
# なぜ機械翻訳が必要か？

（同書三十頁）

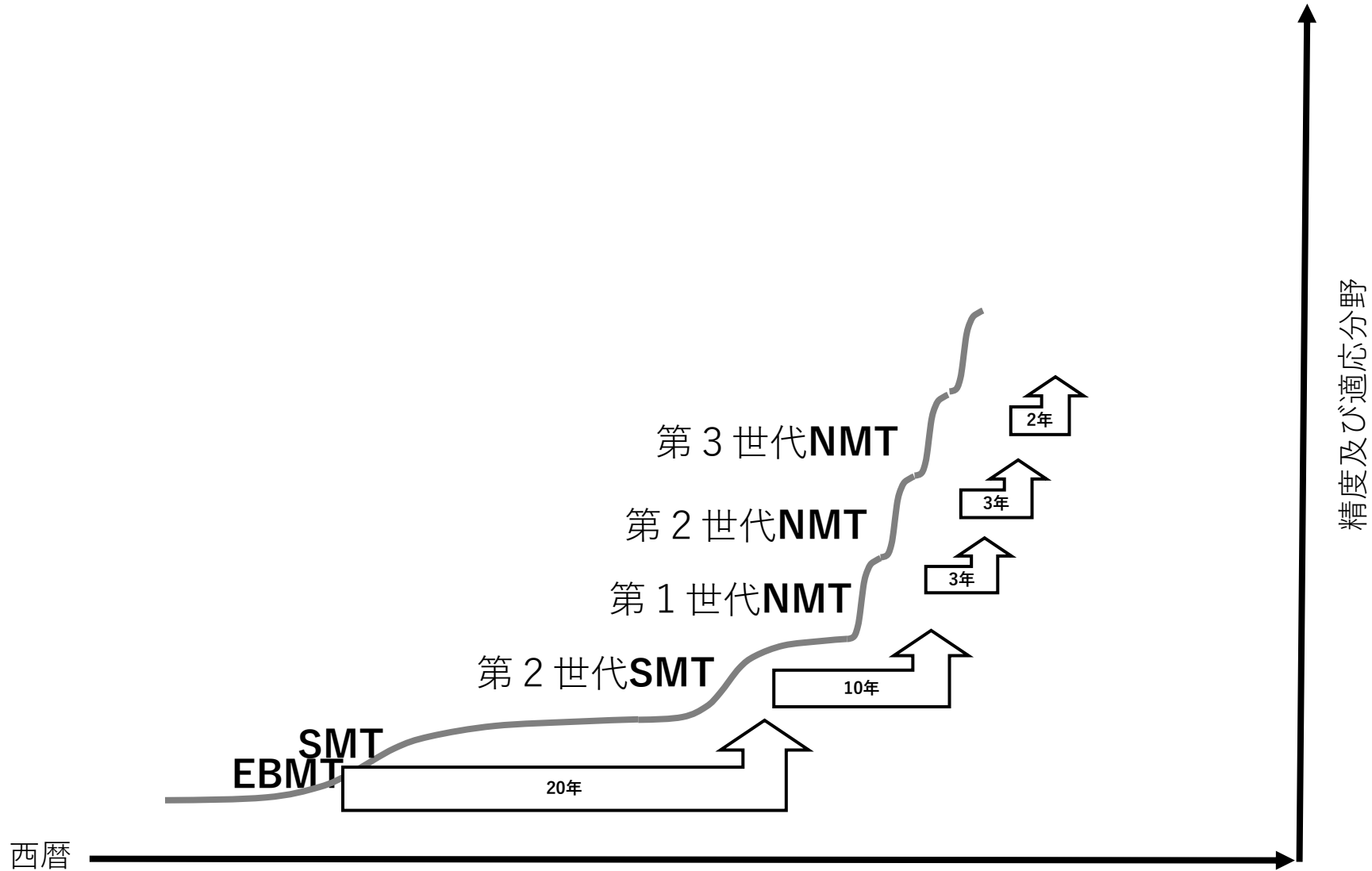
社会の状況はますます翻訳を必要とする時代にはいりつつあった。交通・通信の発展によって世界はせまくなり、貿易がますます世界的規模で行なわれ、多種多様な情報が瞬時に世界をかきめぐるいう時代となってきた。その結果、**翻訳に対する需要は急速に増大し、翻訳者の数がまったく不足する**という状況がおこりはじめていたのである。

翻訳需要  $y$  は  
(5年で2倍ずつ)  
 $x$  回増えた  
として、今は  
1986年の256倍

Exponential Function with Base 2



# コーパスベース技術で破壊的進化が連続



# AI～コーパスベース技術の到達点～

## 対訳データ(原文と翻訳文)

人類は、国家や地域、  
民族や世代など、あ  
らゆる境界を越えて、  
相互の理解を深め、

Humanity has  
achieved progress  
as it has  
deepened its  
mutual

## 学習済みのニューラルネット

0.034524 0.008209 -0.036567 0.000317 0.052561 -0.057137 -0.049069 0.026325 -  
0.006376 0.005100 -0.029406 0.001348 0.007813 0.011046 -0.068254 0.049055  
0.015998 0.050705 -0.032895 -0.057581 0.020686 0.013625 0.006844 -0.005559 -  
0.021837 -0.007766 -0.016833 -0.052483 -0.015757 -0.018395 -0.039223 -  
0.025823 -0.009908 -0.010407 0.029266 0.033712 -0.017647 -0.019884 -0.068798  
0.002268 0.046325 -0.015774 0.004865 -0.017944 -0.007810 -0.020594 -0.004333  
-0.013813 0.037423 0.028352 -0.038967 -0.015373 -0.076406 -0.012378 -  
0.009984 -0.026934 -0.006178 -0.010026 -0.024871 0.021154 0.012995 0.011356

良いデータとは大量のデータだ。

ICT is also the  
infrastructure that  
supports  
humanity's  
advanced  
intellectual and  
economic  
activities.

0.081033 0.012333 0.010833 0.004107 0.034337 0.103478 0.031230 0.071343  
0.068661 0.056756 0.113601 0.020353 -0.035564 -0.114542 -0.074180 -0.005134  
-0.022966 0.027060 0.083860 0.041548 -0.040060 -0.019864 -0.091129 -0.070714  
0.038758 -0.042537 0.053469 0.011317 -0.008002 -0.090652 -0.010616 0.067415  
0.053682 -0.015672 -0.144984 -0.005615 -0.043907 0.070575 0.127576 -0.027835  
0.072007 -0.061721 -0.097822 0.079427 -0.005288 0.050989 -0.091070 0.015273  
0.021031 -0.072851 -0.037229 0.083061 0.128342 -0.044585 -0.137006 0.093918  
-0.033246 -0.008432 0.080802 -0.109352 0.083443 -0.045618 -0.033986 0.027695  
-0.115097 0.223685 -0.083115 -0.032411 -0.040359 0.101720 0.077344 -0.020547  
0.110848 0.103057 0.026371 0.033157 -0.015743 0.058732 0.129566 0.158788  
0.017742 0.067832 -0.016020 0.021265

<http://www.nict.go.jp/およびhttp://www.nict.go.jp/en/index.html>



# AIによる 文書翻訳の 自動化



# 英文ニュースもAI で上手に翻訳で きる

～2023年1月27日付のAPを例に～

<https://apnews.com/article/production-facilities-climate-and-environment-business-d095684168e9f6a2634ee9316007f994>

1. The manufacture of “green steel” moved one step closer to reality Friday as Massachusetts-based Boston Metal announced a \$120 million investment from the world’s second-largest steelmaker, ArcelorMittal.
2. Boston Metal will use the injection of funds to expand production at a pilot plant in Woburn, near Boston, and help launch commercial production in Brazil.
3. The company uses renewable electricity to convert iron ore into steel.
4. Steel is one of the world’s dirtiest heavy industries.
5. Three-quarters of world production uses a traditional method that burns through train loads of coal to heat the furnaces and drive the reaction that releases pure iron from ore.
6. Making steel releases more climate-warming carbon dioxide than any other industry, according to the International Energy Agency — about 8% of worldwide emissions.

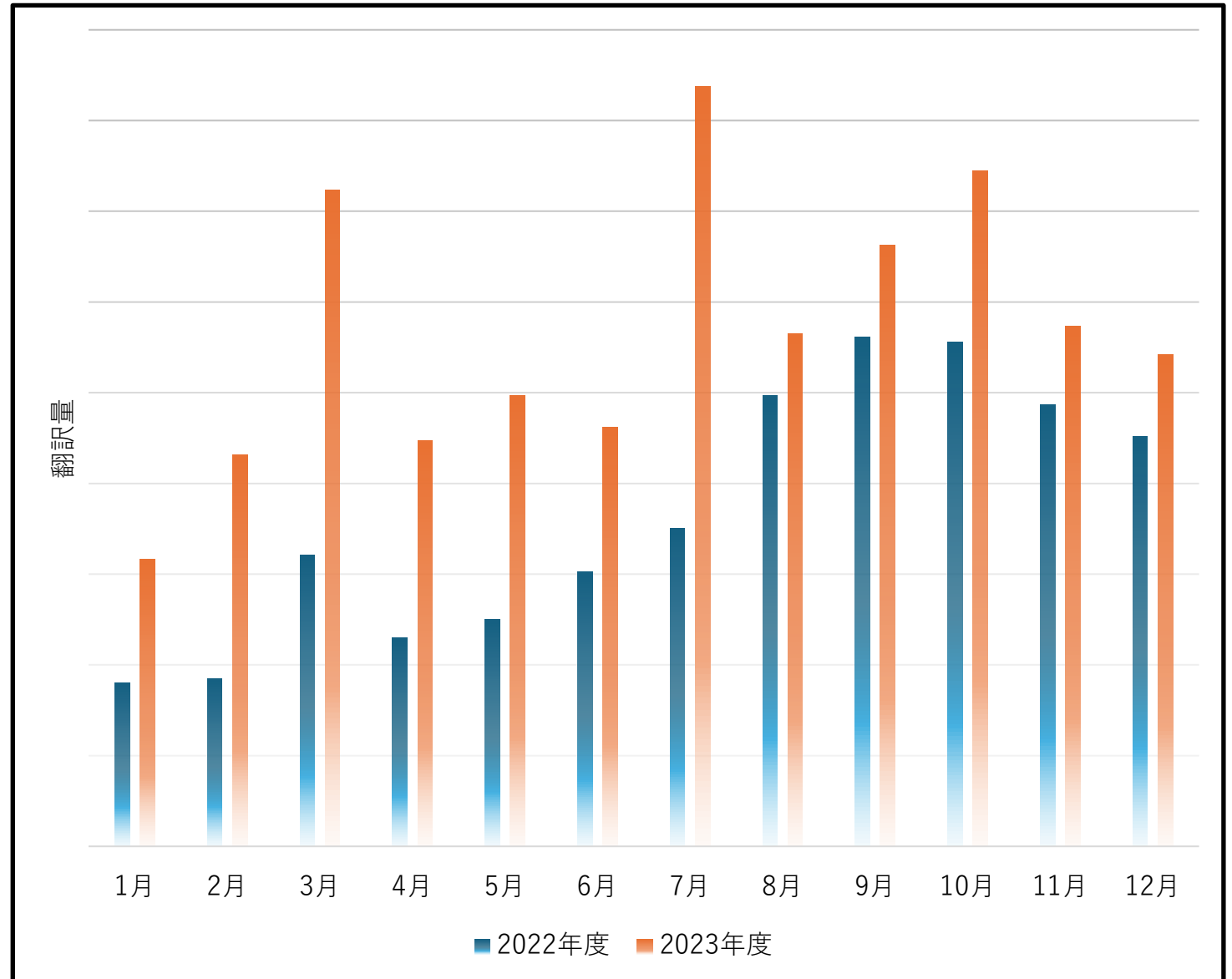
1. マサチューセッツ州に拠点を置く Boston Metal が、世界第2位の鉄鋼メーカーである ArcelorMittal から1億2000万ドルの投資を発表したことで、「グリーン・スチール」の製造は金曜日に現実に一歩近づいた。
2. 調達した資金は、ボストン近郊のウォバーンにあるパイロットプラントでの生産拡大や、ブラジルでの商業生産開始に充てる。
3. 同社は再生可能な電力を使用して鉄鉱石を鉄鋼に変換している。
4. 鉄鋼は世界で最も汚い重工業の一つである。
5. 世界の生産量の4分の3は、列車に積まれた石炭を燃焼させて炉を加熱し、鉱石から純鉄を放出する反応を促進する伝統的な方法を用いている。
6. 国際エネルギー機関(IEA)によると、鉄鋼製造は他のどの産業よりも地球温暖化につながる二酸化炭素の排出量が多く、世界の排出量の約8%を占めている。

# 金融専門用語 に強い金融庁 モデル※

金融庁の利用は顕著に増加

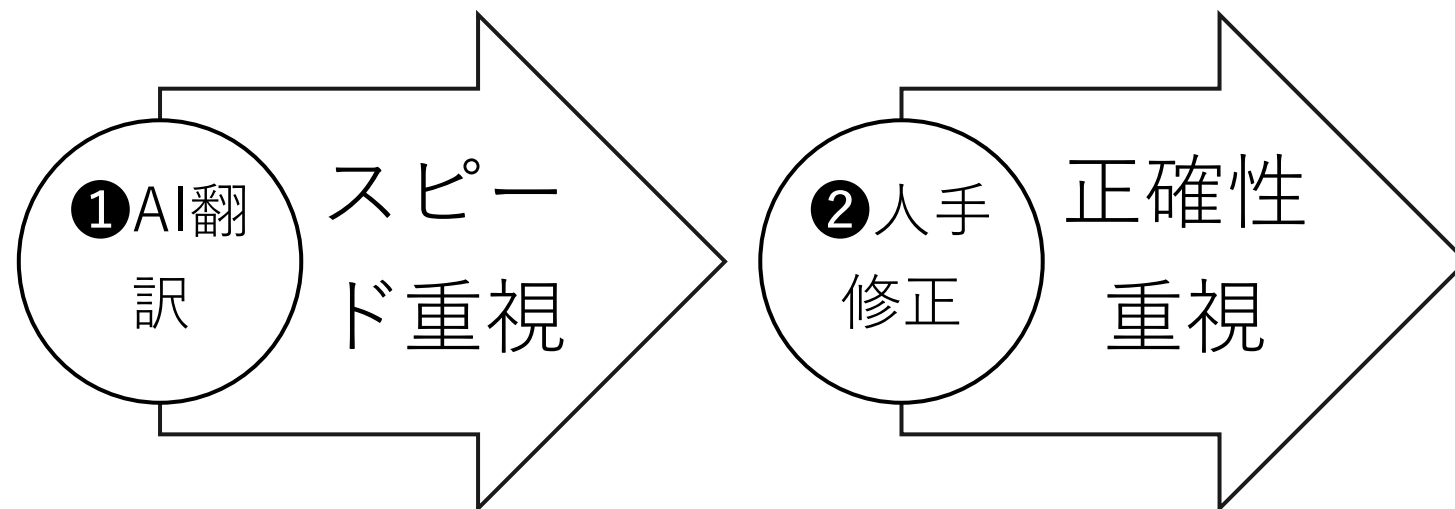
※

<https://www.fsa.go.jp/news/r3/sonota/20220311/20220311.html>



# 法務省による法令の英訳の2段階公開

<https://www.japaneselawtranslation.go.jp/ja>



2024-11-20	行政手続法施行令を暫定版から最終版に差し替えました。
2024-11-20	行政手続法（第15条第1項、第3項及び第4項、第16条第1項、第22条第3項、第31条未施行）を暫定版から最終版に差し替えました。
2024-11-20	商標法を暫定版から最終版に差し替えました。
2024-11-19	特定化学物質の環境への排出量の把握等及び管理の改善の促進に関する法律施行規則（暫定版）を公開しました。
2024-11-19	担保付社債信託法を暫定版から最終版に差し替えました。

研究に必要な英語論文は日本語で読める

<https://mt-auto-minhon-mlt.ucri.jgn-x.jp/>

<https://mt-auto-minhon-mlt.ucri.jgn-x.jp/content/tool/search/>

みんなの自動翻訳 eTexTra<sup>®</sup> Language メニュー

## 論文翻訳検索

ヘルプ

NICT 統計

検索対象  arXiv  PubMed 検索言語  日本語  英語

期間  全期間  過去1週間  過去1か月  過去12か月  過去24か月  過去36か月

<https://arxiv.org/abs/1809.07037> PDF

2018/09/19 (更新日: 2023/10/18) Benjamin Marie, Rui Wang, Atsushi Fujita, Masao Utiyama, and Eiichiro Sumita

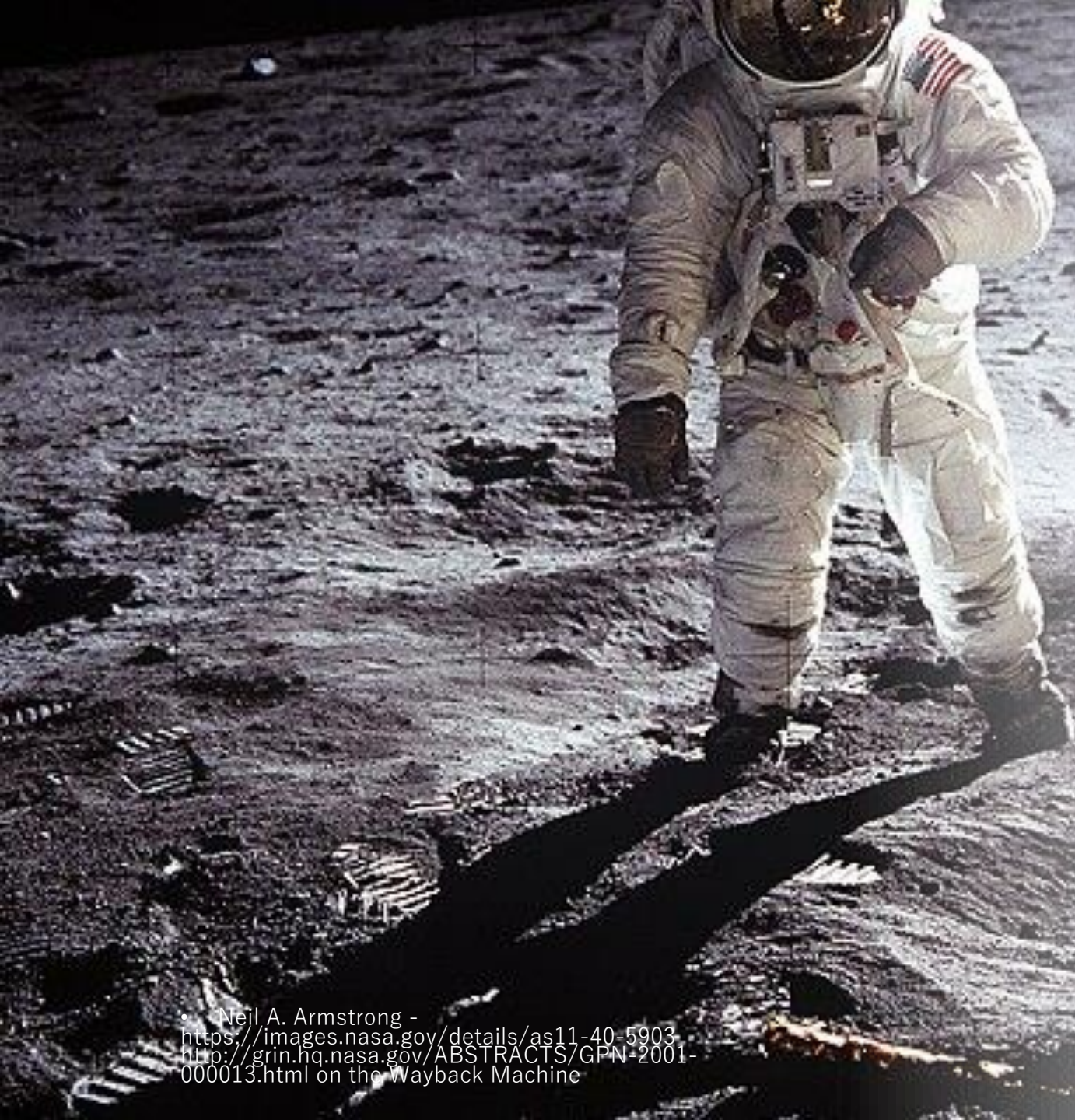
### WMT18ニュース翻訳タスクのためのNICTのニューラルおよび統計機械翻訳システム

本論文では、WMT18共有ニュース翻訳タスクへのNICTの参加について述べる。エストニア語-英語、フィンランド語-英語、トルコ語-英語、中国語-英語の4言語ペアの8つの翻訳方向に参加した。各翻訳方向に対して、最先端の統計(SMT)およびニューラル(NMT)機械翻訳システムを準備した。このNMTシステムは、新しい増分トレーニングフレームワークで生成した大量の逆翻訳された単一言語データで拡大された提供された並列データを使用して、トランスフォーマーアーキテクチャでトレーニングされた。タスクへの主な提案は、SMTシステムとNMTシステムの単純な組み合わせの結果である。このシステムは、BLEUケースに従って、エストニア語-英語とフィンランド語-英語の言語ペア(翻訳)で第1位にランクされている。

### NICT's Neural and Statistical Machine Translation Systems for the WMT18 News Translation Task

This paper presents the NICT's participation to the WMT18 shared news translation task. We participated in the eight translation directions of four language pairs: Estonian-English, Finnish-English, Turkish-English and Chinese-English. For each translation direction, we prepared state-of-the-art statistical (SMT)

英語文献DB(PubMedとarXiv)の日本語検索を提供  
(8月末時点で、arXiv 215万件、PubMedは3,765万件)



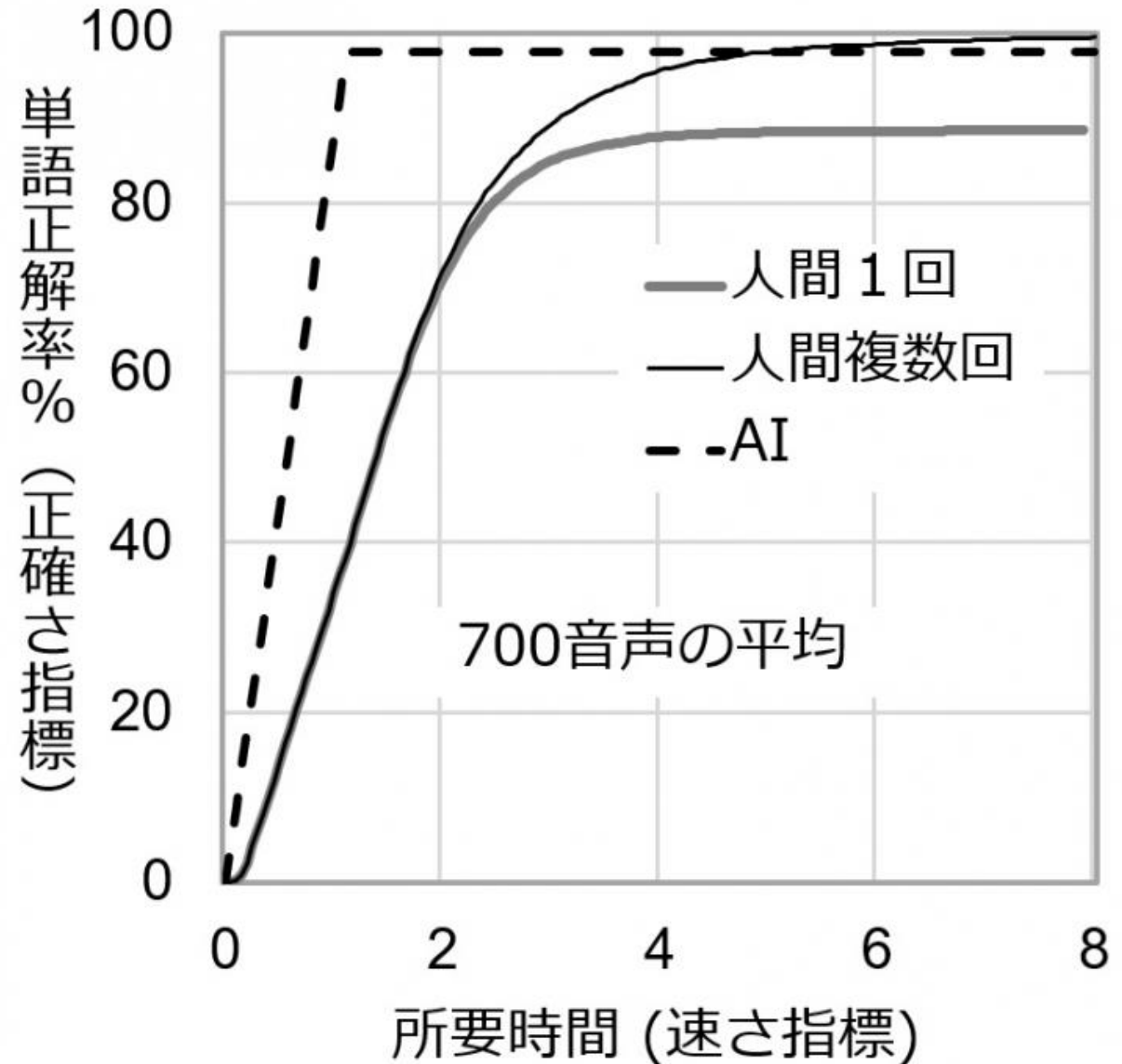
文書の自動  
翻訳は普及  
して水や空  
気のよ  
うになっ  
た

AIによる  
**同時通訳**の  
自動化



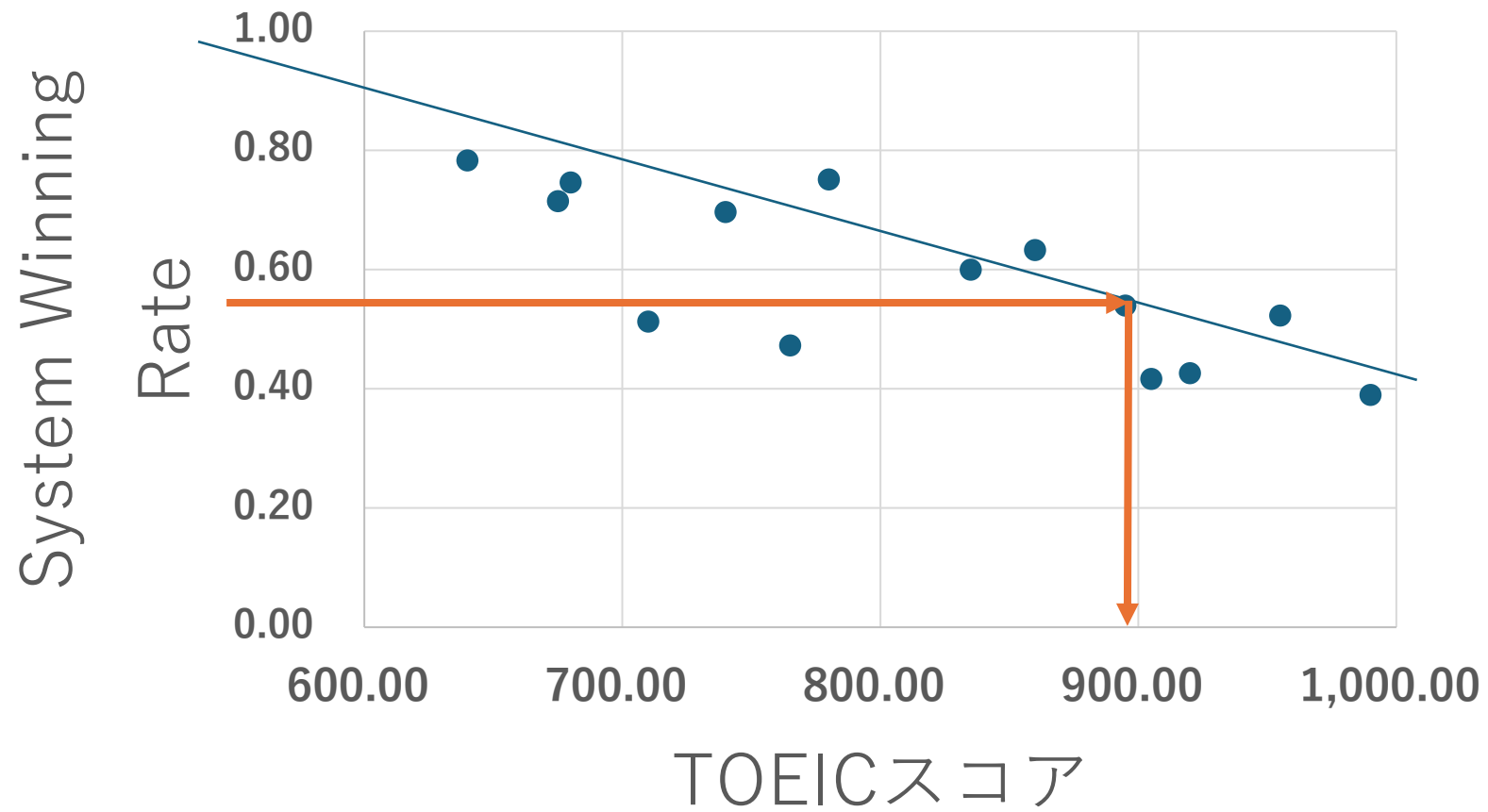
# 進化したAIは 日本語認識で 日本人越え

「AIは人間を超えたのか？音声認識技術を“ガチで”競わせてみた」  
<https://newswitch.jp/p/30140>

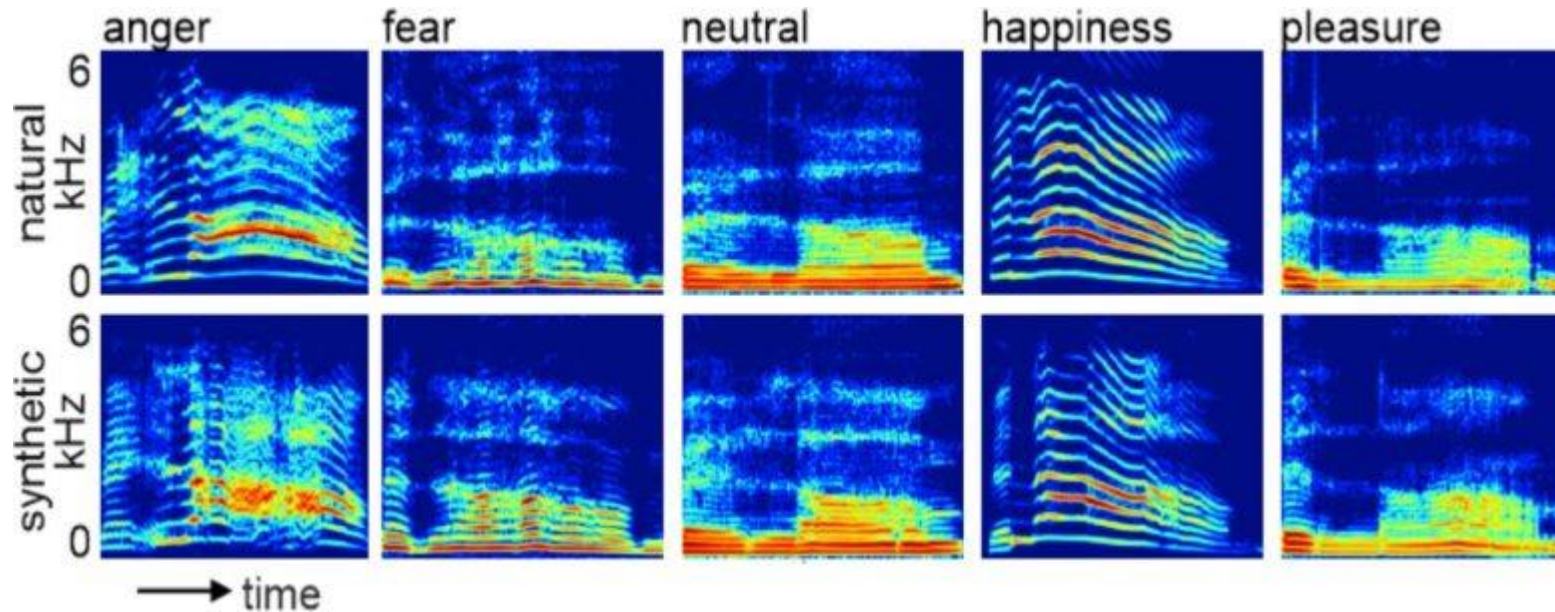




# 2020年の音声翻訳のTOEICスコア



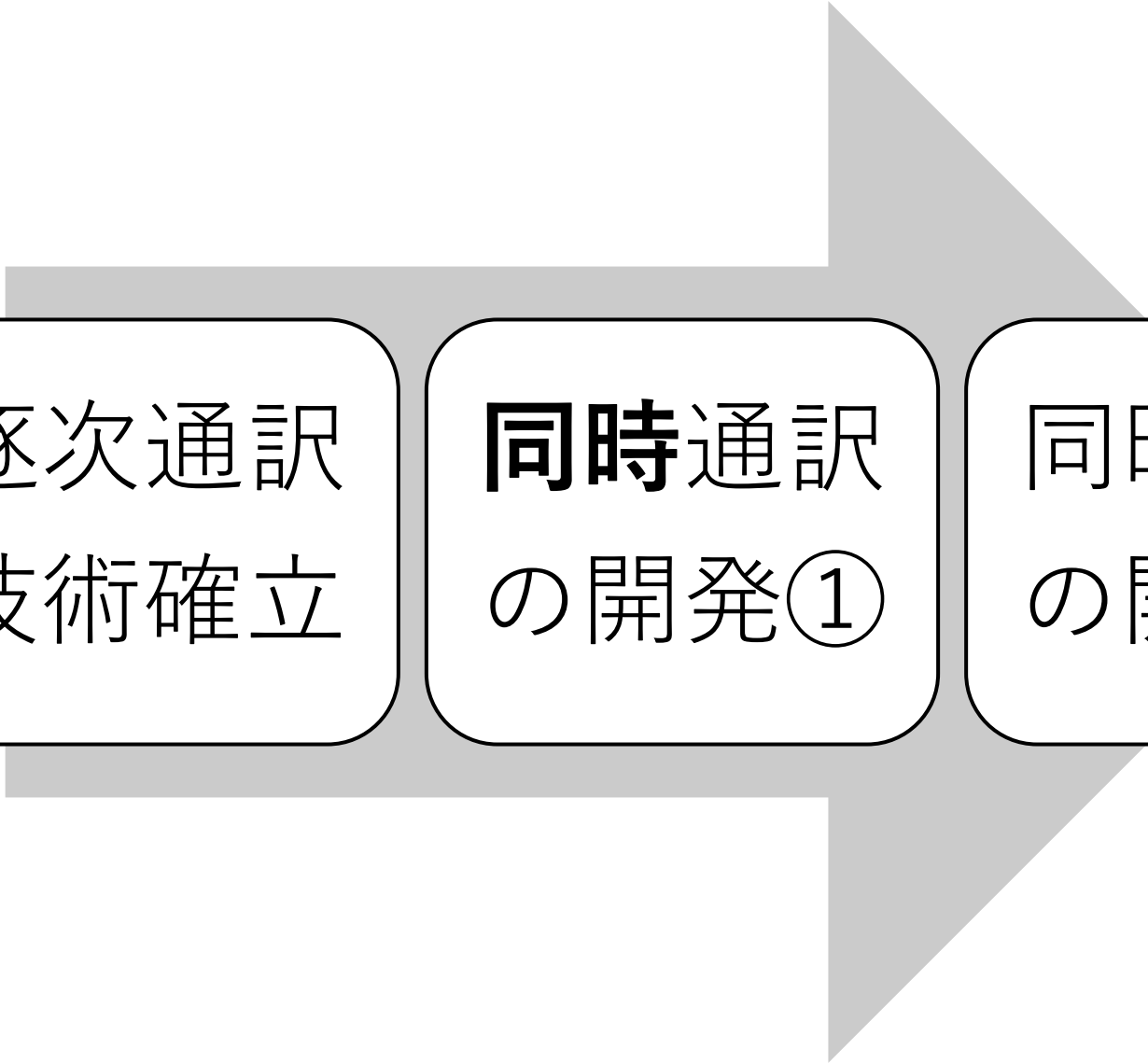
# AIなのか、それとも本物なのか？



二つの区別が大変つきにくいです。

人間の声を正しく識別できたのは56%、AIの声は50.5%だったため、どちらのタイプの声も同じように聞き分けるのが苦手だったことになります。

<https://scitechdaily.com/the-great-voice-trick-is-it-ai-or-is-it-real/>

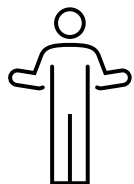


逐次通訳  
技術確立

**同時**通訳  
の開発①

同時通訳  
の開発②

## 同時通訳の分割点データ



英語の分割点データ

I will never forget / the  
time / when we first  
met. /

深層学習

cytonNeuralSentSeg



英語の分割点予測  
ニューラルネット

同時通訳者による  
チャンク分割を模倣

# 利用者を待たせない翻訳単位「チャンク」導入※

原文			①チャンク単位翻訳(声)
文1	チャンク1	前回	Previously
	チャンク2	ちょっと太り気味という 数値が出ていて	the results showed I was a little too fat,
	チャンク3	心配だったんだけど...	and I was concerned.
文2	チャンク4	要注意がなくなって	I didn't receive a warning,
	チャンク5	ほっとしたよ。	and I'm relieved.

「前回」に対して、  
「Previously」の翻  
訳が素早くでる。

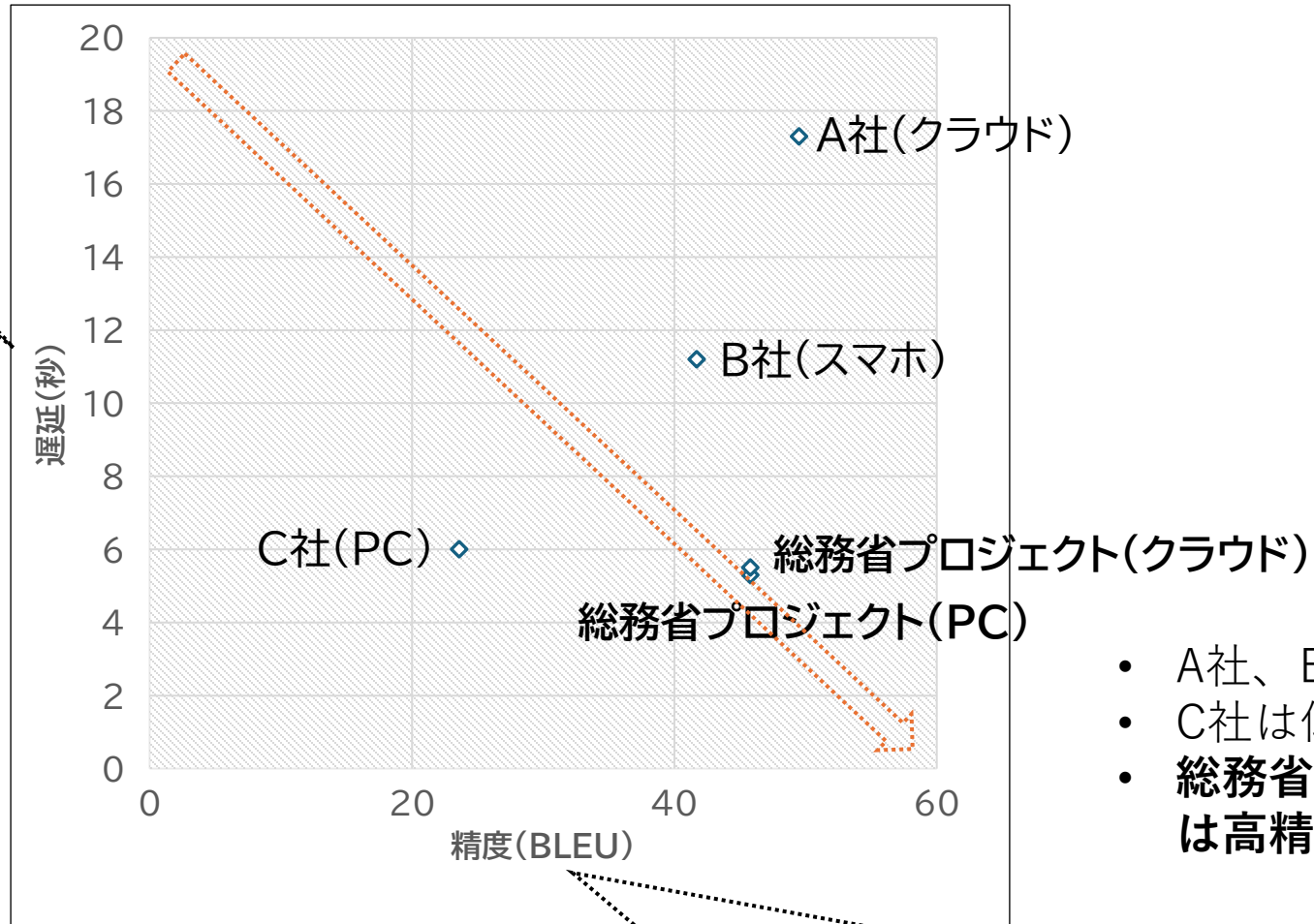
※ チャンクとは、同時通訳者のように短く  
切って翻訳する単位のこと。同時通訳の基  
本特許(特願2021- 27112)  
の出願中

# 日英デモSHORT




# 他社技術との遅延・精度の比較 ～低遅延&高精度～

音声入力から訳出までの遅延時間等)



- A社、B社は高精度だが低遅延でない。
- C社は低遅延だが高精度でない。
- **総務省プロジェクト (PC及びクラウド) は高精度かつ低遅延で、優位性がある。**

BLEU値：翻訳精度を自動評価する標準的指標で大きい方が良い。模範訳とシステム訳の単語列一致で算出される。

A photograph of an astronaut in a white spacesuit standing on the lunar surface. The astronaut is wearing a helmet and has an American flag patch on the right shoulder. The ground is dark and rocky, with a long shadow cast by the astronaut. The background is a bright, hazy sky.

# 同時通訳 も 商品化！

• Neil A. Armstrong -  
<https://images.nasa.gov/details/as11-40-5903>  
<http://grin.hq.nasa.gov/ABSTRACTS/GPN-2001-000013.html> on the Wayback Machine



# AIによる 翻訳/通訳の 『語学』への 影響

Neil A. Armstrong - <https://images.nasa.gov/details/as11-40-5903>  
<http://grin.hq.nasa.gov/ABSTRACTS/GPN-2001-000013.html> on the Wayback Machine

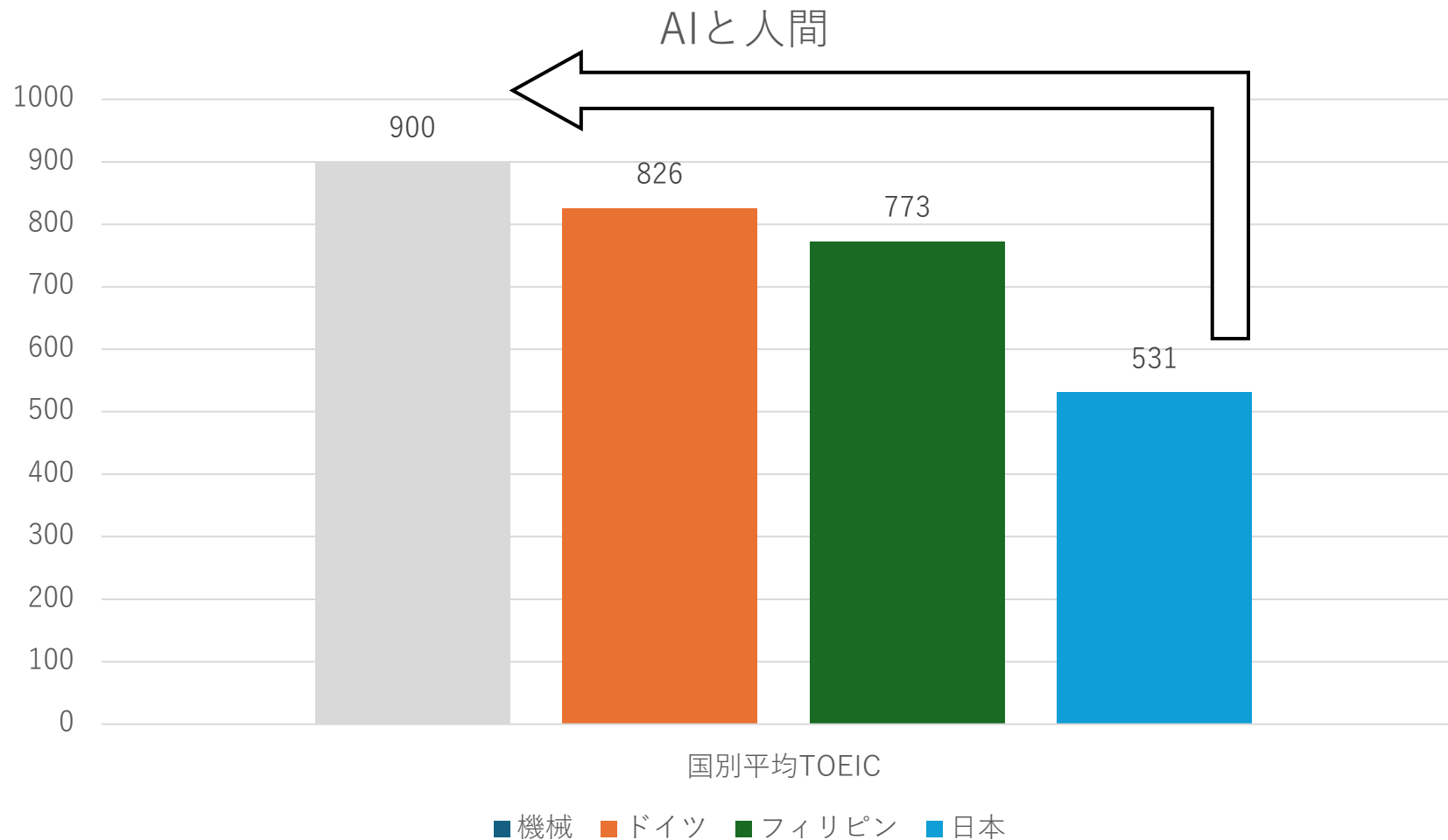
# 教育での パラダイ ムの逆転

～神奈川県厚木市の市立  
小学校成田潤也教諭等の  
実践～

[https://globe.asahi.com/article/14983284?fbclid=IwAR1kxcw\\_BNiJ002ZKfv-L5AHB9Br9nmzF53ve-zxOK-hyoiEj-NOg12kgeo](https://globe.asahi.com/article/14983284?fbclid=IwAR1kxcw_BNiJ002ZKfv-L5AHB9Br9nmzF53ve-zxOK-hyoiEj-NOg12kgeo)

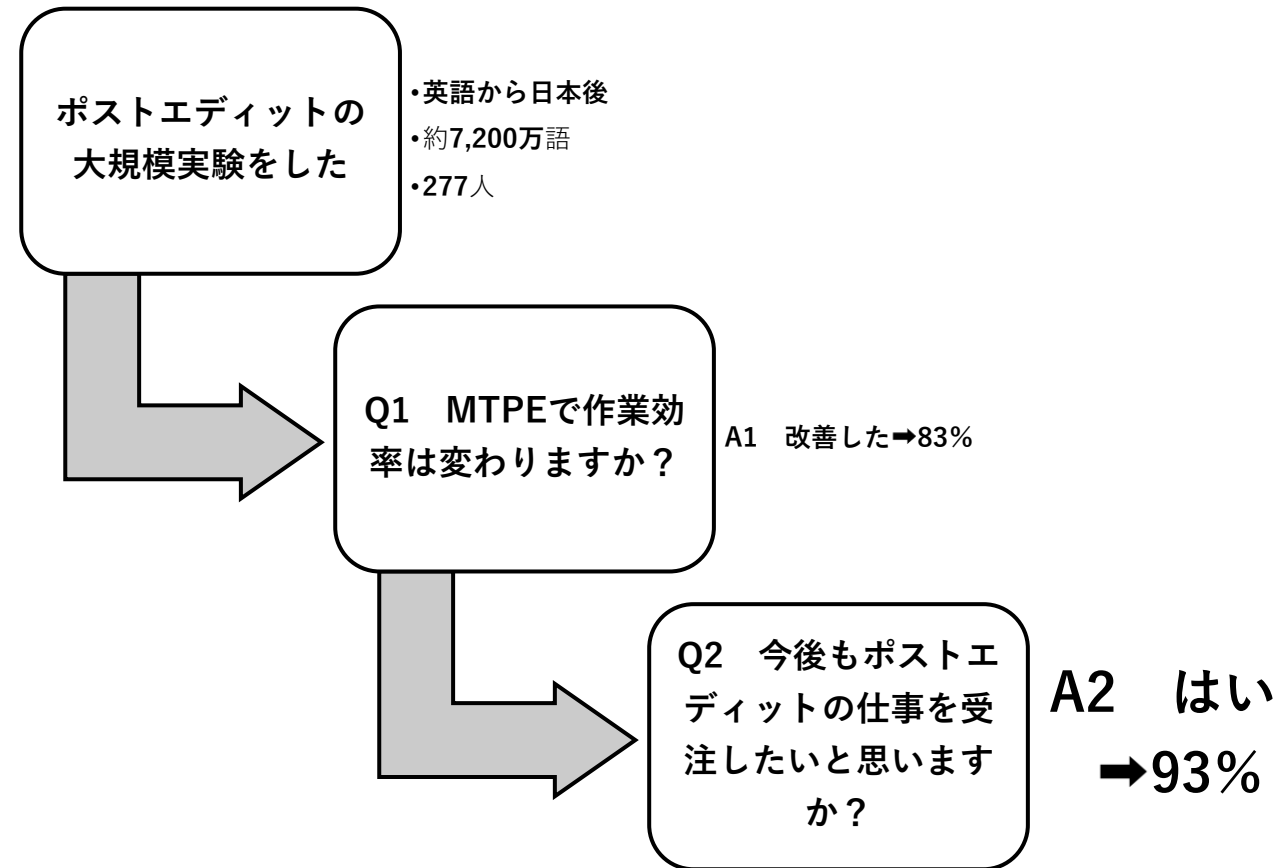


# 進化したAI は一部でSUPER HUMAN



<https://www.iibc-global.org/iibc/press/2021/p174.html>

# 翻訳者、即ちプロも活用へ



A photograph of an astronaut in a white spacesuit standing on the lunar surface. The astronaut is wearing a helmet and has an American flag patch on the chest. The ground is dark and rocky, with a long shadow cast by the astronaut. The background is a bright, hazy sky.

# 語学の 転換点

• Neil A. Armstrong -  
<https://images.nasa.gov/details/as11-40-5903>  
<http://grin.hq.nasa.gov/ABSTRACTS/GPN-2001-000013.html> on the Wayback Machine

# 溫故 知新2

# なぜ機械翻訳が必要か？

（同書百四十一頁）

現在のシステムは翻訳

を一文単位で（中略）

文と文の関係を考えて

翻訳しなければならな

いものは無視している。

# コンテキストが必要①～文脈や視覚情報～

【単数/複数、定/不定】

机の上に本がある➡

- 1) There **is a** book on a table.
- 2) There **are** books on a table.
- 3) There are books on **tables**.



# コンテキストが必要② ~文脈や視覚情報~

【3以上】

教授が学生の中に立っている➡

- 1) The professor stands **between** the students.
- 2) The professor stands **among** the students.

# コンテクストが必要③～ 知識や状況～

**President X will join this party soon.➡**

X**国家主席**はもうすぐこのパーティーに参加します。

X**大統領**はもうすぐこのパーティーに参加します。

X**社長**はもうすぐこのパーティーに参加します。

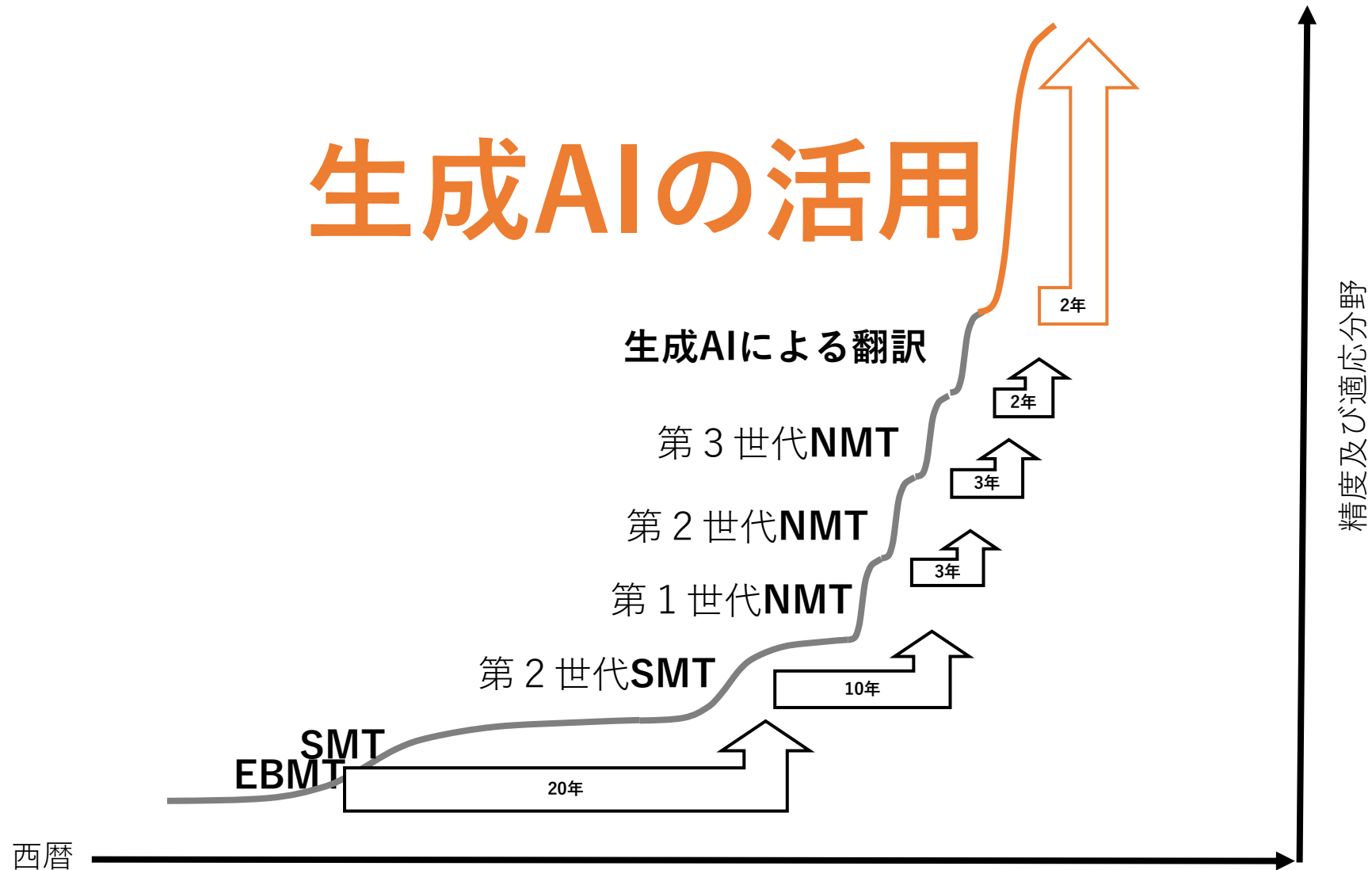
# 否定解釈にコンテクスト必要④～理解～

**All** the questions have **not** been answered.➡

【全否定】すべての質問が答えられていません。

【部分否定】すべての質問が答えられたわけではありません。

# 翻訳・通訳の技術革新が止まらない



# Transformer流の師匠と弟子 (出藍の誉れ)

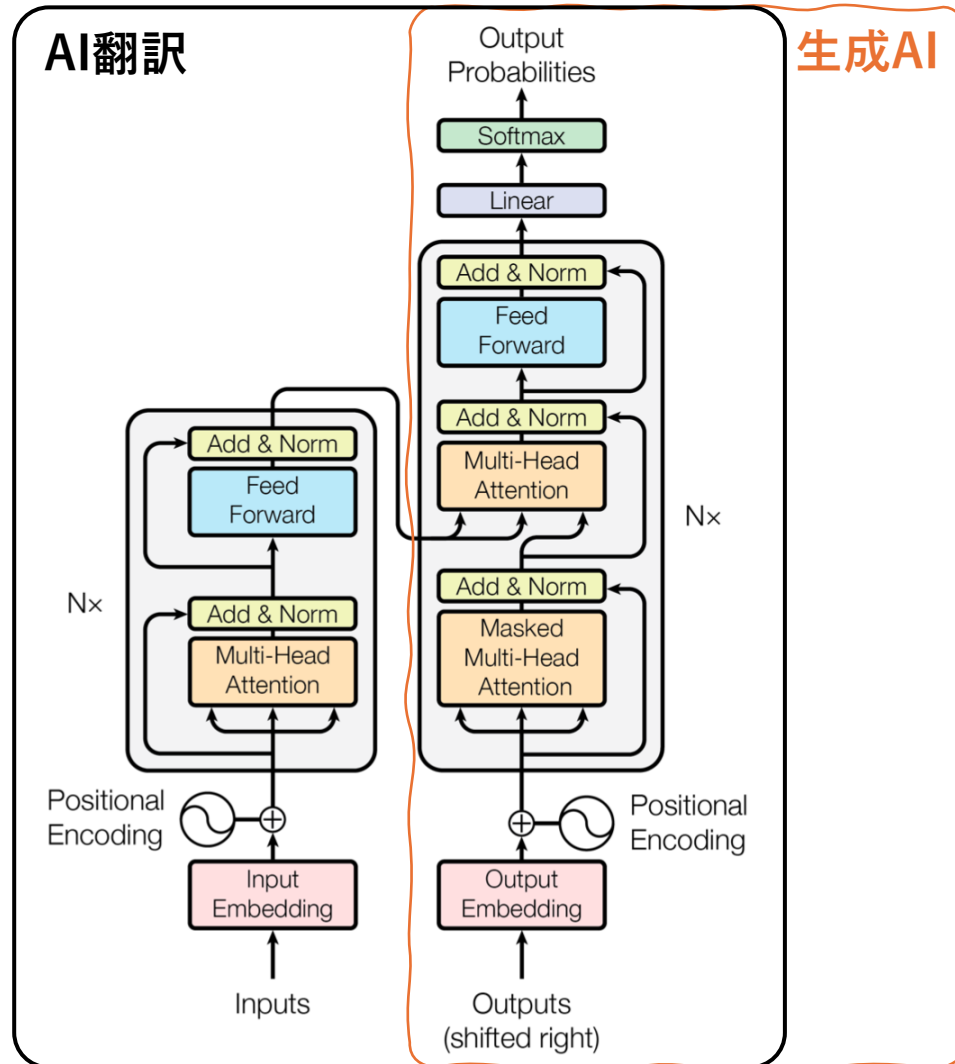


Figure 1: The Transformer - model architecture. (Vaswani 他(2017 ))

# 「たかが暗記というなかれ」の 生成AIの規模感

※ 平均的語数は10万語/  
冊として

	平均的日 本人の読 書量 (12冊/年 *80年)	国会図 書館の 蔵書量 (1千万冊)	議会図 書館の 蔵書量 (4千万 冊)	META社 生成AIの 学習量
単語 数	1億	1兆	4兆	15兆



# 1文を越え たコンテク ストの活用

# コンテキストとは

- 対象の1文に伴う周辺の文、関連文書、背景情報、世界・文化に関する知識、画像情報等の総称。
  - ✓ コンテキストに依存せずに翻訳が確定する場合は従来の自動化法で高精度を達成できた。
  - ✓ 一方、訳文がコンテキストに依存する場合はコンテキスト把握を計算機上に実現する必要があり、従来上手く翻訳が出来なかった。



# 人間の同時通訳者と計算機の従来の同時通訳の差

同時通訳	種別	人間	従来の機械	今後の機械
	利用する情報	<b>全情報</b>	音声の言語情報 (コンテキストなし)	情報を増やす (コンテキストあり)
ゴールにボール を入れる スポーツ	種別	ハンドボール	サッカー	
	利用する部位	<b>手も足も使う</b>	足だけ使う	

# 人間のように発話を通訳

ちょうちょうをおいかけてすごした。

現在の私は解剖医だが昔は昆虫少年だった。ちょうちょう(蝶々)をおいかけてすごした。

➡ I am an anatomist now, but I was once a boy who loved insects. I spent my time chasing **butterflies**..

現在の私は小説家だが昔は新聞記者だった。ちょうちょう(町長)をおいかけてすごした。

➡ I am a novelist now, but I was once a newspaper reporter. I spent my time chasing the **mayor**.

# 漫画翻訳の自動化の近況

2020年 マンガに特化した機械翻訳技術の研究  
開発Mantra社設立<https://mantra.co.jp/>

2021年 マンガ翻訳事業 / 電子マンガストア事  
業のオレンジ社設立<https://orange.inc/ja>

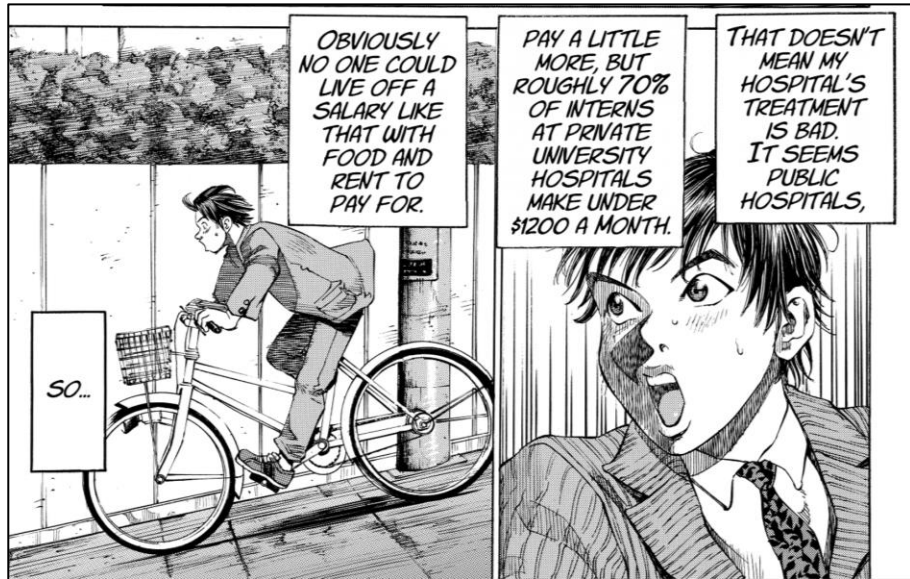
**2023年 「韓国で権威ある翻訳賞、韓国語が“で  
きない”日本人が受賞し物議＝韓国ネット「何の  
問題が？」**

[https://www.recordchina.co.jp/b908920-s39-  
c30-d0191.html](https://www.recordchina.co.jp/b908920-s39-c30-d0191.html)

2024年 サイバーエージェントは、AIを活用し  
漫画などの海外展開を支援する専門組織「AIロー  
カライズセンター」を新設。

2023年2月9日、韓国メディア・韓国経済は「韓国語のできない日本人が人工知能（AI）翻訳機を使ってウェブトゥーン（韓国のウェブ漫画）を翻訳し、韓国で権威ある翻訳賞を受賞する～中略～事態を受け、同院は新人賞の公募制度の改善とAIとの協業範囲について議論することを決めた。～中略～この記事を見た韓国のネットユーザーからは「何の問題があるのか分からない」「パパゴの翻訳文をきれいに整えるのは思いのほか難しい。100%信用することもできない。良い訳が思いつかないときに参考にする程度なら可能だけど、パパゴだけで翻訳賞をとるのは無理」～中略～などの声が寄せられている。

# 「ブラックジャック によろしく 佐藤秀 峰」の（コンテクス ト有り）試訳



1. うちの病院の 待遇が 特別に悪い わけではない → **Our hospital's treatment isn't particularly bad.**
2. 国立ではもう少し 高いらしいが 私大病院の 研修医の月給は 10万円以下のところが 約7割を占める → **I've heard that national hospitals pay a bit more, but about 70% of private hospital resident doctors earn less than 100,000 yen per month.**
3. その給料で 寮も食事もないんだから 生活できるわけない → **With that salary, I don't even get a dorm or meals, so how can I survive?**
4. で—— → **Ugh...**

# 英日小説の(コンテクスト有り)試訳

O・ヘンリーの短編『二十年後』の冒頭を日本語に試訳して比較。コンテクストの有無での訳語選択に差。

原文	従来法	試訳：コンテクスト有り
The policeman on the beat moved up the avenue <b>impressively.</b>	巡回中の警官は <b>印象的に</b> 大通りを進んだ。	巡回中の警官は通りを <b>堂々</b> と歩いていった。

コンテク  
ストで飛  
躍的高精  
度化

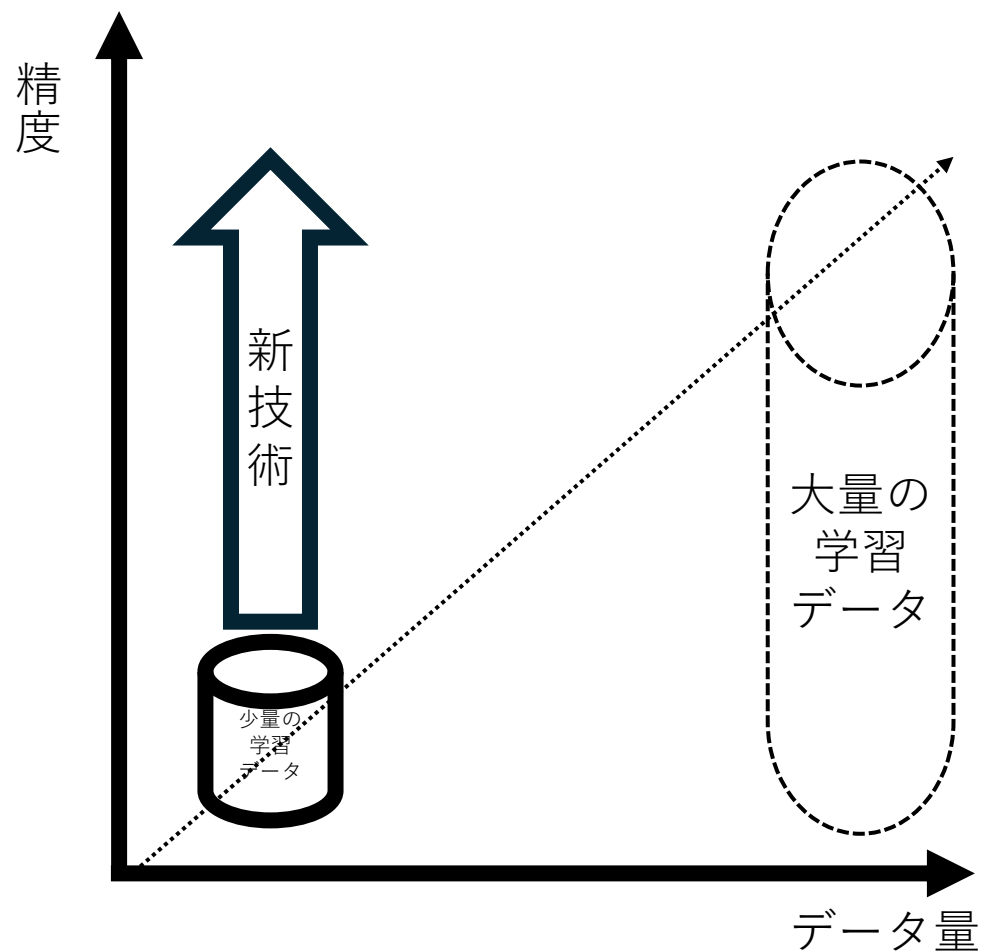




# 爆速的 多言語化

# 小資源での翻訳の高精度化

～「良いデータとは大量のデータだ。」を越える～

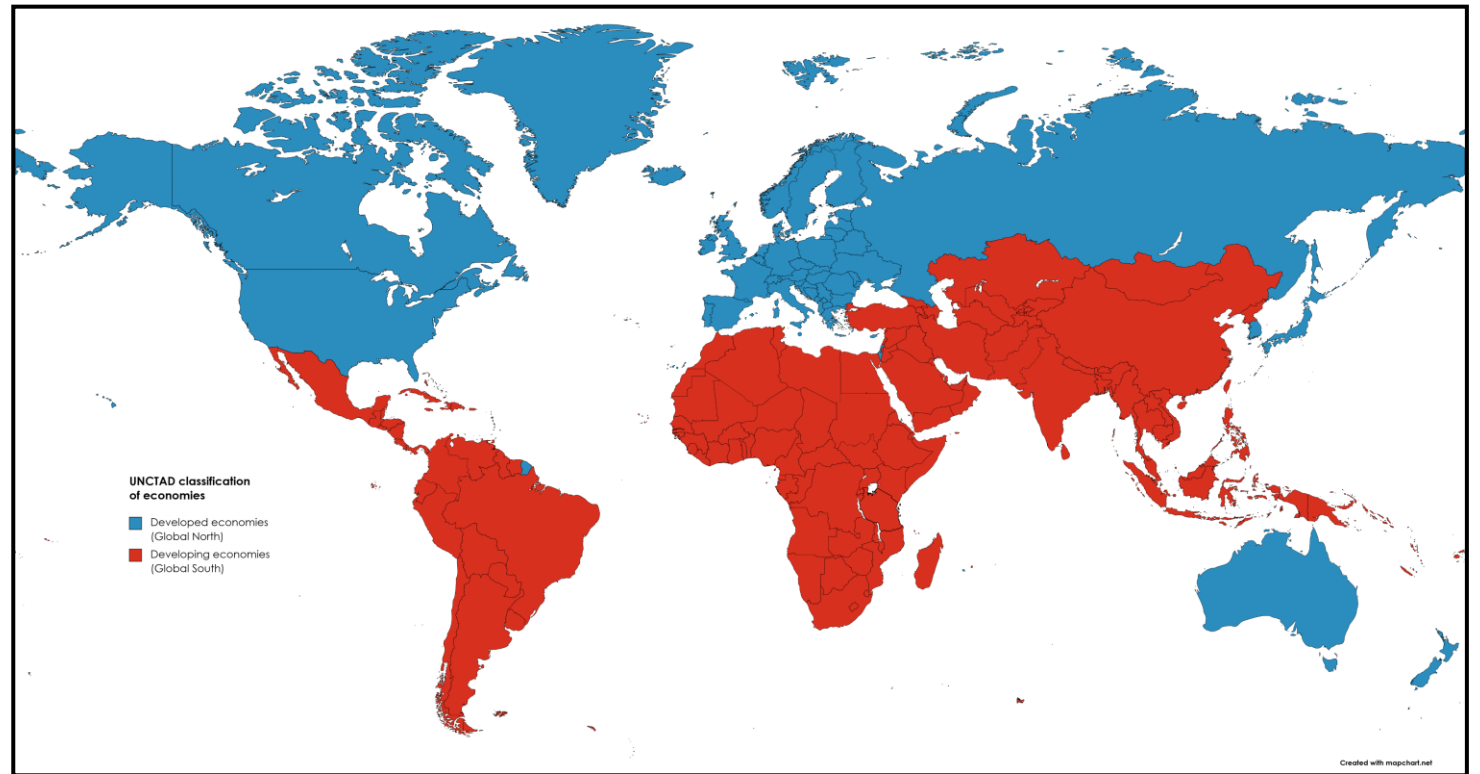




成長を牽引するグローバルサウスとの連携が重要

---

製品/サービスの輸出にはグローバルサウス言語と日本語の多言語コミュニケーションが必須  
(人手は全く足りない)



<https://commons.wikimedia.org/wiki/File:UNCTADclassificationofeconomies.png>

# 自動翻訳で世界の言語が平等に

AI翻訳  
常在

誰もがAIで**意思疎通**でき、  
外国語学習の  
負担は軽減

伸び盛りのグ  
ローバルサウ  
スの**経済圏**と  
も**自由に繋ぐ**

**世界語**の解消

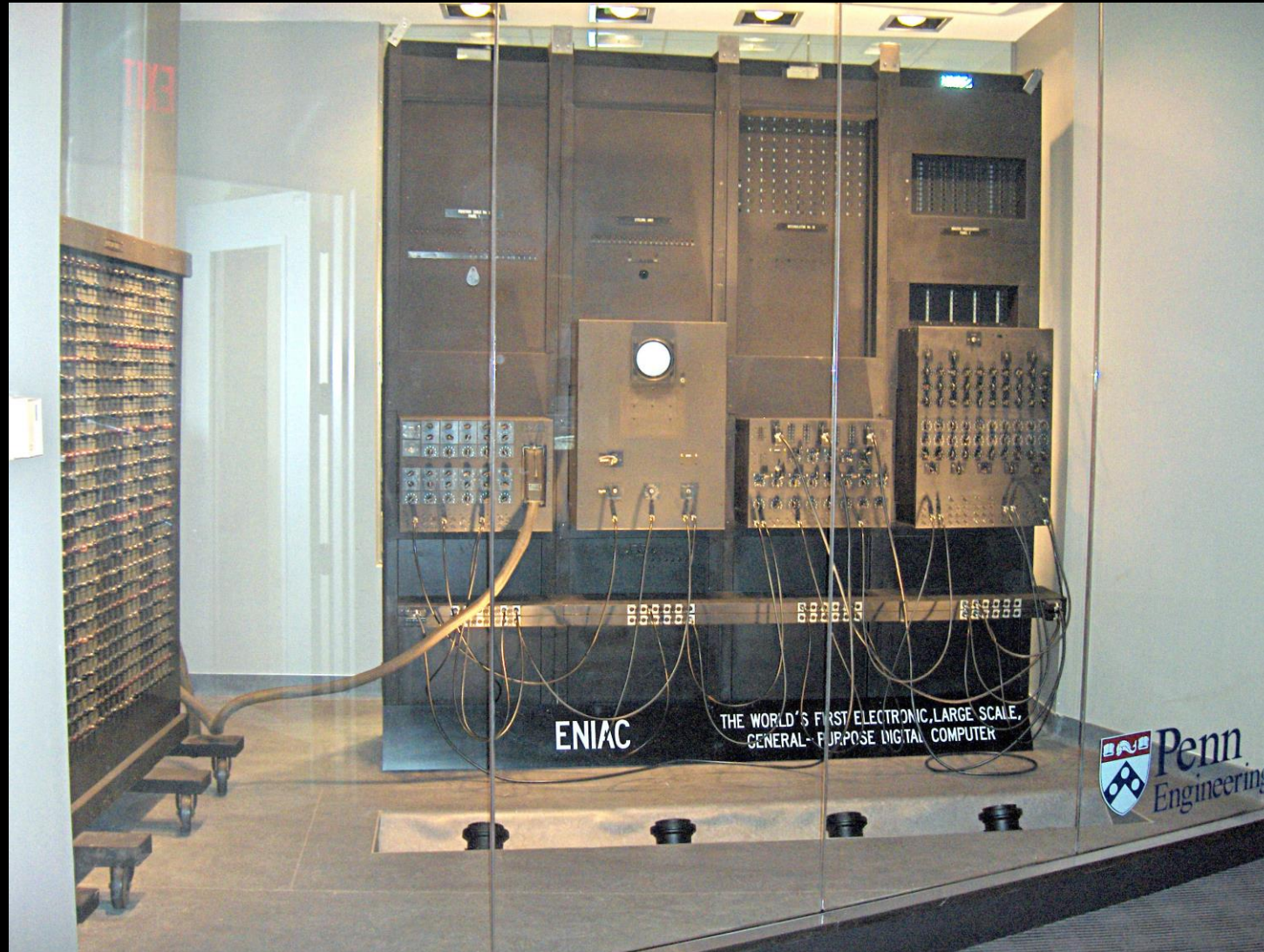
来使  
出言  
訳から多  
訳るから  
翻える  
語化



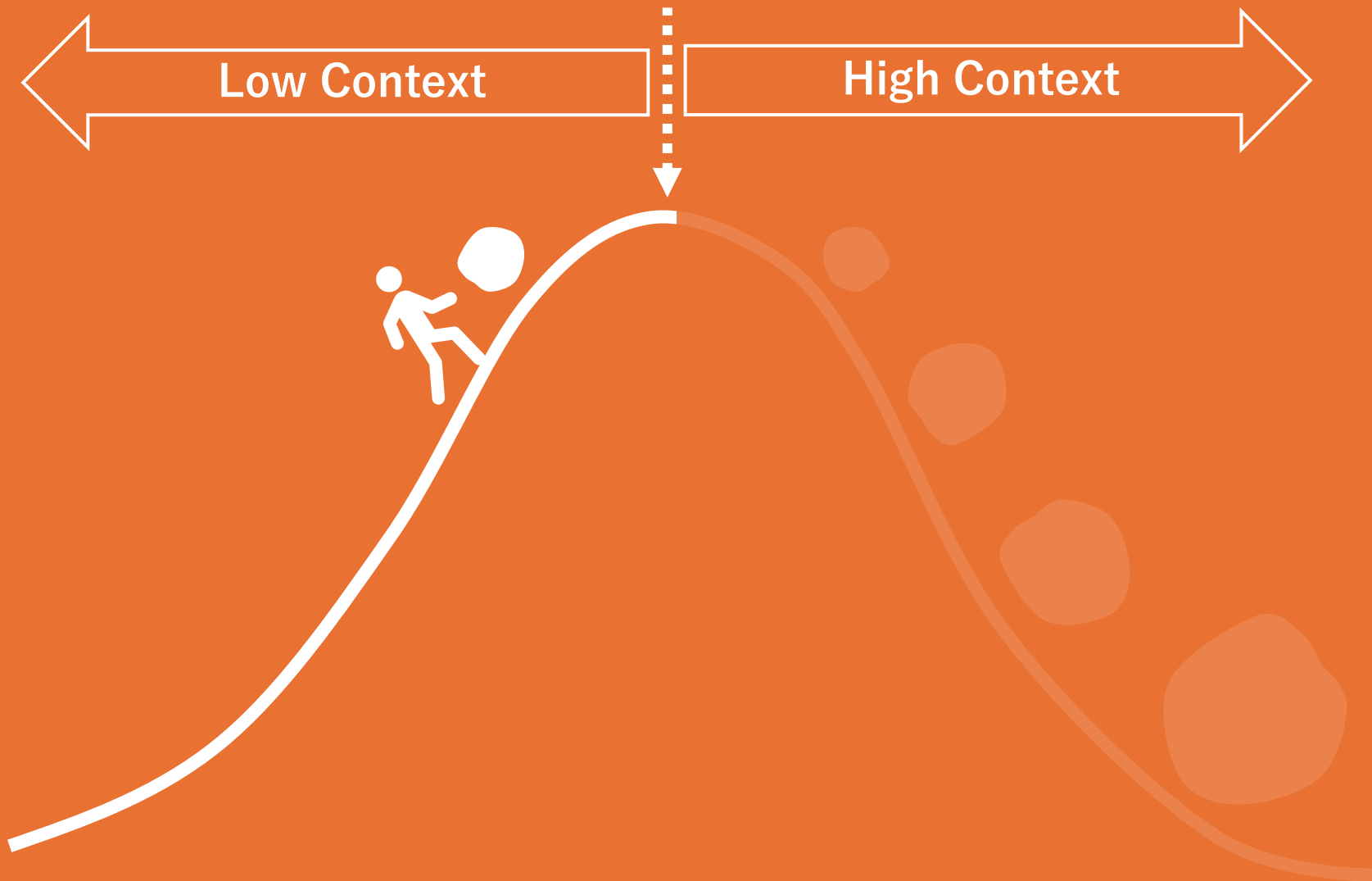


# 翻訳・通訳 の新時代

翻訳の自動化は、第二次世界大戦直後に、  
世界初の汎用計算機の民事転用の一つとして研究開始



# 生成AIのメリットを活用してTipping Pointを越える





**機械翻訳は  
どこまでも  
可能!**