



ービス「ふみのは®」や、2023 年から開始した、一般の方でも手軽に古文書が解読可能なスマホアプリ「古文書カメラ®」を展開しています。

このような背景のもと、この度 TOPPAN は、これまで「くずし字」の解読で培った AI-OCR に関する技術や知見から、中世ギリシャ語の解読が可能な AI-OCR エンジンを開発しました。

## ■ 本 AI-OCR エンジンの特長

### ・中世ギリシャ語の解読

中世ギリシャ語は、時代や書き手によって字形が異なるほか、単語の一部が省略される場合や、現代とは異なる綴りが用いられるなど、表記が一定ではないという特徴があります。また、単語と単語の間に区切りを設けずに文章が書かれていることもあり、専門知識のない現代人には読み取りが困難とされています。本 AI-OCR エンジンでは、100 万字規模の字形や行のデータベースを学習データとして用意することで、中世ギリシャ語文字の解読を実現しました。

### ・ヴァチカン教皇庁図書館の保有データを学習データとして活用

ヴァチカン教皇庁図書館が保有する約 5000 点のギリシャ語写本のうち、既に付加情報を加えた 50 点 (IIIF 画像約 400 枚)と翻刻テキストを AI の学習データとして活用します。写本画像と翻刻テキストの高精度な学習に加え、専門家による目視確認も組み合わせることで、解読精度の向上と品質担保を両立します。これにより、膨大なギリシャ語写本コレクションのテキストデータ化を加速させるとともに、本 AI-OCR エンジンにおける中世ギリシャ語文字の認識精度 95%以上の達成を目指します。

## ■ TOPPAN とヴァチカン教皇庁図書館の取り組みについて

TOPPAN グループは印刷博物館の設立準備以来、1997 年から 29 年間にわたりヴァチカン教皇庁図書館と複数のプロジェクトで協力してきました。グーテンベルク 42 行聖書の高精細デジタルアーカイブ (※3)やキケロ・プロジェクト(※4)、印刷博物館での共同展覧会(※5)など、古文書の解読や文化の継承を推進する取り組みを共同で行っています。2026 年 4 月 25 日には、印刷博物館での共同展覧会として 3 回目の企画展となる「名著誕生展 ヴァチカン教皇庁図書館 III+」を開催します。本展では、ヴァチカン教皇庁図書館より借用した資料とともに、ギリシャ語 OCR のデモンストレーション映像も展示します。

## ■ 今後の目標

TOPPAN グループは、専門知識なしでは解読困難な中世ギリシャ語写本画像に対し、高精度な AI-OCR エンジンにより翻刻文を提供する環境を整備することで、ギリシャ語研究の活性化やさらなる発展に寄与します。また、資料のデジタルアーカイブなどの取り組みを通して、ヴァチカン教皇庁図書館をはじめ、世界各地で収蔵されてきた文化資産の保全と、全世界の人々がアクセス可能な環境整備を両立し、次世代へと継承していくための技術革新を推進していきます。

※1 IIIF (International Image Interoperability Framework)：世界中の図書館・博物館・アーカイブなどでデジタル化された高精細画像を、サイトの枠を超えて相互に利用・共有するための国際的な技術標準・枠組み。この規格に対応することで、異なる機関の画像を一つのビューワ (Mirador など)で並べて比較や、注釈の追加、高度な比較・分析が容易になり、文化財の活用を促進します。

※2 翻刻:古い写本や印刷物に書かれた文章を、現代の楷書体・活字(テキストデータ)に置き換える作業。

※3 グーテンベルク 42 行聖書 高精細デジタルアーカイブ (2000 年)

ヴァチカン教皇庁図書館所蔵のグーテンベルク 42 行聖書 (羊皮紙版、6 葉欠落)を精細デジタル化。また、他館所蔵の資料から抽出したテキスト部分の画像を合成し、欠落していた 6 葉 (12 ページ)をデジタル上で復元。

※4 キケロ・プロジェクト(2005年～現在)

再利用羊皮紙写本(パリンプセスト)が上書きされる前の重要な古代文書を解読・復元するプロジェクト。TOPPAN グループは画像認識・デジタル化技術を活用したスキャナーと解析ソフトウェアを開発。これらにより、可視光画像に紫外線を照射した画像を重ね合わせることで、消失した過去のテキストを抽出・解読。2022年時点で、4万ページを超える画像を読み取り、その内、88冊分の解析画像がヴァチカン教皇庁図書館のWebサイトで公開されています。

※5 印刷博物館での共同展覧会

過去2回(2001年、2015年)開催。

\* 本ニュースリリースに記載された商品・サービス名は各社の商標または登録商標です。

\* 本ニュースリリースに記載された内容は発表日現在のものです。その後予告なしに変更されることがあります。

以 上