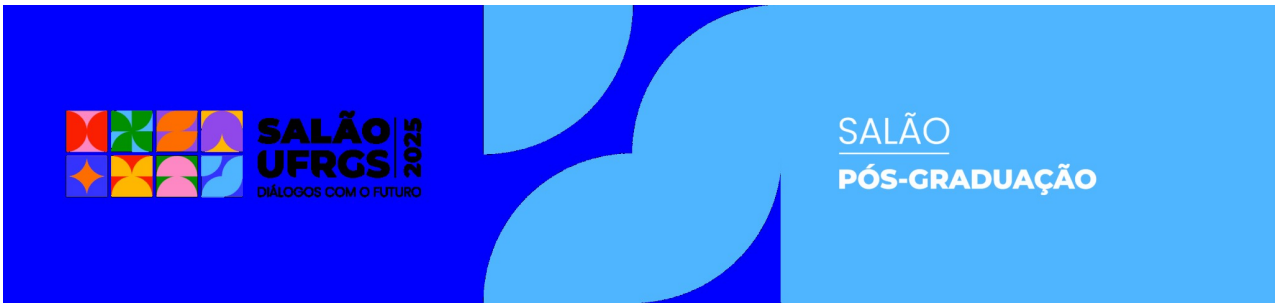


Evento	Salão UFRGS 2025: I SALÃO DA PÓS-GRADUAÇÃO
Ano	2025
Local	Campus do Vale - UFRGS
Título	Handwriting Recognition aplicado aos acervos arquivísticos com a plataforma Transkribus
Autor	LEONARDO PORTO DE BITTENCOURT PEREIRA
Orientador	MOISES ROCKEMBACH

RESUMO: A pesquisa busca abordar as áreas da Arquivologia, Ciência da Informação e Humanidades Digitais de forma interdisciplinar, com foco na incorporação de tecnologias digitais no auxílio à transcrição de documentos históricos. Nesse sentido, destaca-se o uso da Inteligência Artificial, especialmente a partir da tecnologia de Reconhecimento de Texto Manuscrito (HTR), e o papel da plataforma Transkribus nesse processo. A proposta central é refletir criticamente sobre as potencialidades dessa ferramenta no contexto das áreas citadas e da Paleografia Digital. Assim como, explorar as possibilidades de uso da tecnologia HTR e da plataforma Transkribus no ensino. Adentrando mais a fundo no contexto da IA e nas possibilidades de aplicação a serviço de acervos arquivísticos, o Handwriting Recognition (HTR) ou reconhecimento de texto manuscrito, é uma tecnologia que busca compreender, por meio de Optical Character Recognition (OCR), textos manuscritos, transpondo-os para o digital e transformando-os em editáveis. Tal ferramenta possibilita que uma massa documental extensa possa ser transcrita com mais agilidade, ajudando na difusão do seu conteúdo informacional e, por consequência, dos acervos arquivísticos. Diversos fatores podem interferir para o melhor uso do HTR. Alguns deles são: a qualidade da imagem que contém o texto original e a variação na forma que o manuscrito foi feito, já que o mesmo trata-se de caligrafia, e portanto pode variar. A tecnologia HTR é usada por diversas plataformas ao redor do mundo, como a Google Cloud Vision, a Microsoft Azure Computer Vision, entre outras. Assim, o Transkribus emerge como uma plataforma que não só possibilita, mas também fomenta a captura e troca de modelos de transcrição, de modo a otimizar as transcrições paleográficas em todo o mundo. A aplicação desta técnica e a sua otimização em diferentes conjuntos de documentos de diversas origens (década, país, idioma) e culturas é fomentada. Portanto, a utilização desta tecnologia HTR tende a impactar de forma significativa o trabalho de arquivistas, historiadores e pesquisadores de áreas afins, que trabalham ou se interessam por acervos documentais. O objetivo da pesquisa é avaliar a eficácia da tecnologia de Handwriting Text Recognition (HTR) aplicada ao Transkribus, com auxílio do Chat GPT para correções, no contexto da Paleografia Digital. Serão selecionados documentos manuscritos históricos de diferentes períodos e graus de legibilidade. Os critérios de seleção incluem: datação e tipologia do documento; idioma e variações paleográficas; qualidade da digitalização; relevância para a pesquisa paleográfica. Os documentos selecionados serão inseridos no Transkribus para a realização das transcrições automáticas. As etapas incluem: digitalização e upload dos documentos; aplicação da tecnologia de HTR para gerar transcrições automáticas; exportação dos resultados para análise comparativa. Após a transcrição automática, será feita uma análise comparativa com o texto original do documento. Para aprimorar as transcrições, serão utilizadas ferramentas de IA, como o Chat GPT, para sugestões de correção. Serão considerados os seguintes critérios de avaliação: taxa de erro nas transcrições automáticas; tempo necessário para a revisão manual; qualidade das sugestões fornecidas pelo Chat GPT; precisão final da transcrição após revisão assistida. A pesquisa também será aplicada em um ambiente acadêmico com alunos, para analisar a interação entre humanos e IA na paleografia digital. O processo envolverá: distribuição de documentos transcritos automaticamente; correção manual assistida pelo Chat GPT; registro das percepções dos alunos sobre o uso da tecnologia; comparação dos resultados entre transcrições manuais, automáticas e revisadas por IA. Os dados coletados serão analisados a partir de métricas quantitativas (taxa de erro, tempo de transcrição) e qualitativas (percepção dos alunos). O objetivo é verificar a eficiência da tecnologia HTR e do Chat GPT no auxílio à paleografia digital, bem como seu possível impacto na otimização da gestão de acervos arquivísticos. A presente pesquisa seguirá princípios éticos, garantindo o respeito à integridade dos documentos históricos e a privacidade dos participantes. Será assegurada a

devida autorização para o uso de documentos e a obtenção do consentimento informado dos alunos envolvidos.

PALAVRAS-CHAVE: Inteligência Artificial; Reconhecimento de Texto Manuscrito (HTR); Transkribus.