



آرشیوها و هوش مصنوعی: مروری بر مباحث جاری و چشم‌اندازهای آینده^۱

جووانی کولایتزا^۲، توبیاس بلانکه^۳، چارلز یورگز^۴ و جولیا نوردگراف^۵
ترجمه: سید محمد مهدی طبیبیان^۶، امیررضا اصنافی^۷

تحول رقمی، آرشیوهای قدیمی و جدید را به داده تبدیل می‌کند. در نتیجه، خودکارسازی در قالب فنون هوش مصنوعی، بیش‌ازپیش، هم برای گسترش مقیاس فعالیت‌های سنتی نگهداری اسناد و هم برای آزمودن شیوه‌های نوین ثبت، سازمان‌دهی و دسترسی به اسناد به‌کار گرفته می‌شود. در این مقاله، تحولات اخیر در نقطه تلاقی هوش مصنوعی با اندیشه و عمل آرشیوی را بررسی می‌کنیم. مرور ما بر این بدنه روبه‌رشد ادبیات، براساس مدل پیوستار اسناد سازمان‌دهی شده است. در ادبیات مربوط به آرشیوها و هوش مصنوعی، چهار مضمون کلی را شناسایی می‌کنیم: ملاحظات نظری و حرفه‌ای، خودکارسازی فرایندهای نگهداری اسناد، سازمان‌دهی و دسترسی به آرشیوها و شکل‌های نوین آرشیوهای رقمی. در پایان، روندهای نوظهور و جهت‌گیری‌های آینده پژوهش را برجسته می‌کنیم. این جهت‌گیری‌ها شامل به‌کارگیری اصول نگهداری اسناد درباره خود داده‌ها و فرایندهایی است که هوش مصنوعی نوین بر آنها متکی است و نیز ادغام ساختاریافته‌تر. هرچند همراه با آگاهی انتقادی. هوش مصنوعی در سامانه‌ها و رویه‌های آرشیوی است.

کلیدواژه‌ها:
آرشیو،
نگهداری اسناد،
یادگیری ماشینی،
هوش مصنوعی

۱. این مقاله ترجمه‌ای است از:
Colavizza, G., Blanke, T.,
Jeurgens, C., & Noordegraaf,
J. (2021). "Archives and AI: An
overview of current debates
and future perspectives."

ACM Journal on Computing
and Cultural Heritage,
15(1), Article 4. <https://doi.org/10.1145/3479010>

۲. دانشکده علوم انسانی،
دانشگاه آمستردام، آمستردام،
هلند؛ g.colavizza@uva.nl
۳. اداره مرکزی دانشگاه، دانشگاه
آمستردام، آمستردام، هلند؛
t.blanke@uva.nl

۴. دانشکده علوم انسانی، دانشگاه
آمستردام، آمستردام، هلند؛
K.J.P.F.M.Jeurgens@uva.nl

۵. دانشکده علوم انسانی،
دانشگاه آمستردام، آمستردام،
هلند؛ J.J.Noordegraaf@uva.nl

۶. دانشجوی کارشناسی
ارشد، گروه علم اطلاعات و
دانش‌شناسی، دانشگاه شهید
بهشتی، تهران، ایران؛
tabibiyan@gmail.com

۷. دانشیار، گروه علم اطلاعات
و دانش‌شناسی، دانشگاه شهید
بهشتی، تهران، ایران؛
aasnaifi@gmail.com

استناد: کولایتزا، جووانی، بلانکه، توبیاس، یورگز، چارلز و نوردگراف، جولیا. (۱۴۰۵). آرشیوها و هوش
مصنوعی: مروری بر مباحث جاری و چشم‌اندازهای آینده. (محمد مهدی طبیبیان و امیررضا اصنافی،
مترجمان). *آرشیو ملی*، ۹ (۱)، ۴-۲۸ (اثر اصلی در سال ۲۰۲۱ منتشر شده است).

۱. مقدمه مترجمان

با ظهور «الگوریتم‌های یادگیری عمیق»^۱ و در سال‌های اخیر، «مدل‌های زبانی بزرگ»^۲، پارادایم آرشیوهای رقمی در حال گذار از اتکای انحصاری به رویکردهای سنتی مبتنی بر نمایه‌سازی کلیدواژه‌ای به بهره‌گیری از «بازنمایی‌های بُرداری»^۳ و «فضاهای نهفته»^۴ است. این تحول، با بهره‌گیری از مدل‌های یادگیری ماشینی، ظرفیت تحلیل معنایی و استخراج دانش ضمنی از حجم عظیم داده‌های غیرساختارمند را افزایش داده است. همچنین، رویکردهایی مانند «تولید تقویت‌شده با بازایی»^۵ و «گراف‌های دانش»^۶ می‌توانند به برقراری پیوندهای هستی‌شناختی میان اسناد ناهمگون و پشتیبانی از برخی استنتاج‌های مبتنی بر داده کمک کنند. با این حال، این تحولات را نمی‌توان به معنای تبدیل کامل آرشیوها به سامانه‌های «هوشمند و خودمختار» یا محوشدن قطعی مرز میان ذخیره‌سازی داده و تولید دانش در بستر حافظه رقمی تاریخی دانست؛ بلکه این ظرفیت‌ها، در کنار فرصت‌های تازه، پرسش‌هایی جدی درباره نقش آرشیودار، ارزشیابی، اعتماد و نحوه ادغام مسئولانه هوش مصنوعی در رویه‌های آرشیوی مطرح می‌کنند. از این منظر، انتخاب و ترجمه مقاله حاضر اهمیت می‌یابد؛ زیرا نویسندگان با بررسی نقطه تلاقی هوش مصنوعی و اندیشه و عمل آرشیوی، تحولات این حوزه را در چارچوب مدل پیوستار اسناد تحلیل می‌کنند و ظرفیت‌ها، محدودیت‌ها و دلالت‌های حرفه‌ای کاربرد هوش مصنوعی در آرشیوها را به بحث می‌گذارند. هرچند این مقاله پیش از گسترش فراگیر مدل‌های زبانی بزرگ منتشر شده است، مرور منسجم مباحث و چارچوب مفهومی آن می‌تواند مبنایی برای فهم تحولات جدیدتر و ارزیابی انتقادی آینده رابطه آرشیوها و هوش مصنوعی فراهم کند.

۲. مقدمه

مدت‌ها پیش از آن که مفهوم «داده‌های بزرگ»^۷ مطرح شود، آرشیوها مجموعه‌های خود را با کیلومترها پرونده و پوشه می‌سنجیدند. تلاش‌های گسترده رقمی‌سازی در سراسر جهان تاکنون دست‌کم بخشی از این مجموعه‌ها را به داده‌های رقمی تبدیل کرده‌اند. علاوه بر این، از دهه ۱۹۹۰ به بعد، دولت‌ها و دیگر نهادهای مرتبط با امور آرشیوی، به‌طور فزاینده‌ای فعالیت‌های خود را به‌صورت رقمی انجام داده‌اند. این تغییر بلافاصله به تحول در رویه‌ها و جریان‌های کاری آرشیوی نینجامید. تا زمانی که حجم مجموعه‌ها و اسناد، چنین امکانی را فراهم می‌کرد، فرایند آرشیوی همچنان عمدتاً بر ارزشیابی، انتخاب و بازبینی دستی استوار بود. با این حال، فرصت تداوم این شیوه کار به سرعت رو به پایان است. روزبه‌روز، مجموعه‌های آرشیوی بیشتری رقمی می‌شوند و «اسناد زاده‌شده رقمی»^۸ نیز در مقیاسی فزاینده به آرشیوها سپرده می‌شوند. این وضعیت، امکان‌پذیری انجام دستی فرایندهای آرشیوی را به‌طور فزاینده‌ای کاهش می‌دهد. در عین حال، همچنان باید اسناد را ارزشیابی کرد تا کیفیت و اعتماد، که بنیان آرشیوها را تشکیل می‌دهند، تضمین شوند. از این رو، آرشیوداران انسانی برای کار با داده‌های بزرگ آرشیوی به پشتیبانی عامل‌های ماشینی نیاز دارند.

1. Deep Learning Algorithms
2. Large Language Models (LLMs)
3. Vector Representations
4. Latent Spaces
5. Retrieval-Augmented Generation (RAG)
6. Knowledge Graphs
7. Big Data
8. Born-Digital Records

بدین ترتیب، نقش آرشیوداران دگرگون می‌شود؛ زیرا باید بیاموزند چگونه از استدلال ماشینی در ارزشیابی و انتخاب بهره بگیرند و هم‌زمان، ارزیابی‌های ماشین‌ها را نیز بسنجند. آرشیو به سازمانی مبتنی بر داده‌های بزرگ تبدیل می‌شود و مانند دیگر سازمان‌های داده‌های بزرگ، برای مواجهه با این تحول ناگزیر است دست‌کم تاحدی به هوش مصنوعی - عمدتاً در قالب یادگیری ماشینی - اتکا کند. در سراسر جهان، آرشیوها ظرفیت‌های هوش مصنوعی را برای سازمان‌دهی جریان‌های کاری خود متناسب با داده‌های بزرگ توسعه می‌دهند و داده‌های بزرگ خود را نیز در اختیار سازمان‌های بیرونی قرار می‌دهند. ده سال پیش، فعالیت‌های مرتبط با هوش مصنوعی در آرشیوها عمدتاً جنبه آزمایشی داشت و ظرفیت‌های این فناوری را نشان می‌داد؛ زیرا راه‌های جدیدی برای کار با بخش‌های مشخصی از مجموعه‌های آرشیوی، مانند مجموعه‌های روزنامه‌های رقمی، فراهم می‌کرد. این وضعیت تغییر کرده است. در سال‌های اخیر، روند تازه‌ای مشاهده کرده‌ایم که طی آن هوش مصنوعی در سراسر فرایندهای نگهداری اسناد - که ماهیت آرشیوها را شکل می‌دهند - به کار گرفته می‌شود. در این مقاله، پژوهش‌های اخیر در نقطه تلاقی هوش مصنوعی و فرایندهای آرشیوی را با بهره‌گیری از «مدل پیوستار اسناد»^۱ به عنوان چارچوب تحلیلی بررسی می‌کنیم.

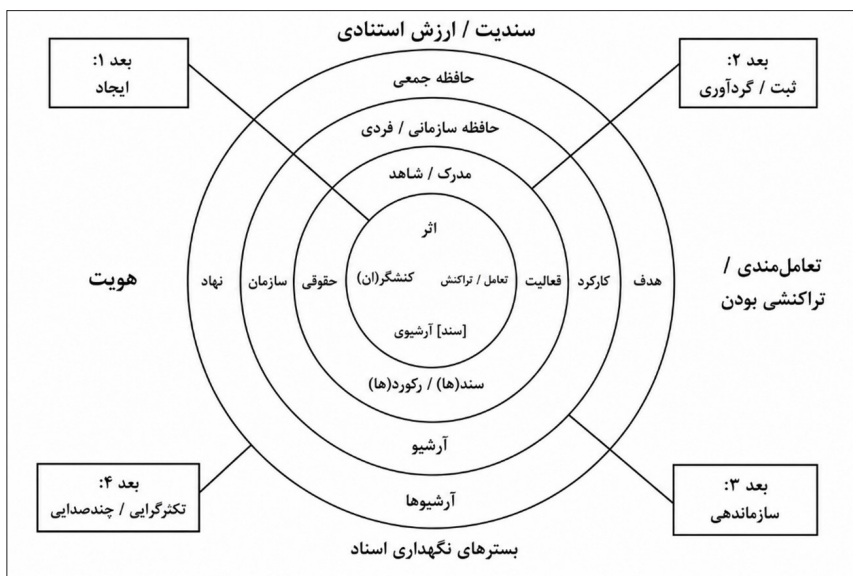
تا آنجا که می‌دانیم، تاکنون چنین مرور جامعی درباره رابطه میان هوش مصنوعی و آرشیوها ارائه نشده است. یک پژوهش، مروری مفصل بر حوزه مرتبط تاریخ رقمی ارائه می‌دهد (Romein et al., 2020)، درحالی‌که پژوهشی دیگر، مروری نزدیک به موضوع حاضر درباره میراث فرهنگی به طور کلی عرضه می‌کند (Fiorucci et al., 2020). اگرچه هیچ‌یک از این دو مرور، پیوستار اسناد را که از مفاهیم تعریف‌کننده آرشیوهاست، در نظر نمی‌گیرند، هر دو نشان می‌دهند که یادگیری ماشینی در بخش میراث فرهنگی، در قالب آزمایش‌های پراکنده متعددی که بر ویژگی‌های مجموعه‌های منفرد تمرکز دارند، به طور گسترده به کار گرفته شده و اکنون در مؤسسات حافظه رواج یافته است. در پژوهشی دیگر، چگونگی تأثیر یادگیری ماشینی بر شیوه‌های آرشیوی از طریق حوزه فنی پردازش خودکار اسناد بررسی شده است (Binmakhshen & Mahmoud, 2020)، اما این بررسی دیدگاه نگهداری اسناد را اتخاذ نمی‌کند. پژوهشی که ارتباط نزدیک‌تری با موضوع مورد نظر ما دارد، اثر مارسیانو^۲ و همکاران است (Marciano et al., 2018). نویسندگان این پژوهش نیز از این فرض آغاز می‌کنند که شیوه‌های آرشیوی با روش‌های نوین و پیشرفته رقمی، مانند یادگیری ماشینی، دگرگون خواهند شد. آنان با بررسی مجموعه‌ای از مطالعات موردی، حوزه میان‌رشته‌ای جدیدی را در نقطه تلاقی آرشیو و علوم رایانه توصیف می‌کنند و پیشنهاد‌های تفصیلی برای تغییر در آموزش آرشیوی ارائه می‌دهند. مقاله ما به این ترتیب سازمان‌دهی شده است: ابتدا دامنه بررسی و روش‌شناسی گردآوری آثار مرتبط را روشن می‌کنیم. سپس، ادبیات را بررسی و در چهار دسته موضوعی کلی که شناسایی کرده‌ایم، سازمان‌دهی می‌کنیم. در نهایت، این روندها را به طور انتقادی به بحث می‌گذاریم و فرصت‌های آتی این حوزه مطالعاتی را برجسته می‌کنیم.

1. Records Continuum Model
2. Marciano

۳. حوزه بررسی

نگارش مقاله‌ای مروری درباره چگونگی بحث در مجلات و متون علمی پیرامون تأثیر هوش مصنوعی بر آرشیوهای مؤسسه‌ای و فرایندهای گسترده‌تر آرشیوی در جامعه، مستلزم تعیین دقیق حوزه مطالعه است. توجه ما تنها به رابطه میان هوش مصنوعی و آرشیوهای موجود در مجموعه‌های مؤسسات میراثی محدود نمی‌شود، بلکه رابطه میان هوش مصنوعی و اسناد را در تمامی ابعاد ایجاد، مدیریت و استفاده آن‌ها بررسی می‌کنیم. به همین دلیل، رویکردی جامع به نگهداری اسناد اتخاذ می‌کنیم که حوزه‌هایی را دربر می‌گیرد که به طور سنتی در مدیریت اسناد و آرشیوازی یکدیگر متمایز بوده‌اند. فعالیت‌های نگهداری اسناد با هدف ایجاد، حفظ و فراهم کردن دسترسی به اسناد کامل، دقیق و قابل اعتماد تراکنش‌ها، در قالب اطلاعات ثبت شده، انجام می‌شوند. نگهداری اسناد بر اصول اساسی‌ای مانند اصالت، زمینه و سندیت استوار است.

این دیدگاه جامع در زمینه نگهداری اسناد به وضوح در نظریه پیوستار اسناد پذیرفته شده است (Gilliland, 2014; Upward, 2018). در مرور ادبیاتی که به هوش مصنوعی در حوزه نگهداری اسناد می‌پردازد، بررسی می‌کنیم که این موضوع چگونه در چهار بُعد مدل پیوستار اسناد که در شکل شماره ۱ نشان داده شده است، بازتاب می‌یابد.



شکل ۱. مدل پیوستار اسناد (Upward, 2018)

مدل پیوستار اسناد، چارچوبی مفید برای ساختاربندی این پژوهش و بحث درباره یافته‌ها فراهم می‌کند. این مدل، چارچوبی کلی و تکررگرا ارائه می‌دهد که تمامی جنبه‌ها و ابعاد آرشیو،

یعنی ایجاد، ثبت، سازمان‌دهی و تکرار اسناد را دربر می‌گیرد. این مدل نه تنها اسناد، بلکه بازیگران و فرایندهایی را نیز شامل می‌شود که در ایجاد، مدیریت، سازمان‌دهی و استفاده اسناد نقش دارند و بر آن‌ها تأثیر می‌گذارند. علاوه بر این، مدل پیوستار اسناد این ابعاد را نه مرحله منفصل، بلکه فرایندهایی چرخه‌ای و متقابلاً تأثیرگذار در نظر می‌گیرد. این ابعاد، نقطه عزیمت مفیدی برای ساختار بندی مطالعات شناسایی شده درباره هوش مصنوعی و آرشیو بوده‌اند؛ زیرا از یک سو، به ما امکان می‌دهند دقیقاً مشخص کنیم که هوش مصنوعی در کدام بخش‌های فرایند آرشیوی به کار گرفته می‌شود و از سوی دیگر، روشن سازیم که کاربرد هوش مصنوعی چگونه بر فرایندهای ایجاد، پردازش و استفاده مجدد از اسناد تأثیر می‌گذارد. در مرور ما، مدل پیوستار اسناد نه تنها برای دسته‌بندی مطالعات مفید بوده، بلکه چارچوبی برای تأمل درباره یافته‌ها نیز فراهم کرده است.

مدل پیوستار اسناد، مدلی خطی و صرفاً مرتبط با چرخه عمر اسناد نیست. ارزش افزوده این مدل در فهم چگونگی تأثیر متقابل ابعاد مختلف آن نهفته است. استفاده مجدد از اسناد برای اهداف جدید ممکن است به بازآفرینی آن‌ها و کشف معانی تازه در امتداد محورهای اثبات‌پذیری، هویت، تراکنش‌مندی یا محمل‌های نگهداری اسناد بینجامد. مدل پیوستار اسناد همچنین به فهم این موضوع کمک می‌کند که فعالیت‌ها و نیازهای ناشی از یک بُعد چگونه ممکن است بر سایر ابعاد تأثیر بگذارند؛ برای مثال، انتظارات و هنجارهای اجتماعی بیرونی در بُعد تکرار چگونه می‌توانند بر جنبه‌های ایجاد، ثبت و سازمان‌دهی اسناد اثرگذار باشند.

در این مرور، اصطلاح هوش مصنوعی را عمدتاً برای اشاره به «یادگیری ماشینی»^۱ به کار می‌بریم؛ حوزه‌ای که به مطالعه و توسعه برنامه‌های رایانه‌ای می‌پردازد که به‌طور خودکار از داده‌ها یاد می‌گیرند. اگرچه می‌دانیم که به‌طور کلی، یادگیری ماشینی را می‌توان یکی از کاربردهای هوش مصنوعی دانست، آن را مؤثرترین و مرتبط‌ترین حوزه هوش مصنوعی برای نگهداری اسناد، براساس تعریف پیش گفته، در نظر می‌گیریم. یادگیری ماشینی همچنین از طریق دیگر حوزه‌های پژوهشی فنی، مانند «پردازش زبان طبیعی»^۲، «بینایی ماشین»^۳ و رباتیک، با نگهداری اسناد ارتباط می‌یابد. علاوه بر این، از اصطلاح هوش مصنوعی استفاده می‌کنیم تا روشن سازیم که توجه ما صرفاً، یا حتی عمدتاً، معطوف به سامانه‌های فنی نیست؛ بلکه کاربرد این سامانه‌ها در نگهداری اسناد. در معنای گسترده آن. و پیامدهای ناشی از این کاربرد را نیز بررسی می‌کنیم. همچنین، اصطلاح هوش مصنوعی را برای در برگرفتن پیامدهای حرفه‌ای، فرهنگی و اجتماعی سامانه‌های خودکار در فرایندهای نگهداری اسناد و برای آرشیو داران به کار می‌بریم.

۳-۱. روش‌شناسی

برای گردآوری نظام‌مند منابع این پژوهش، به شرح زیر عمل کردیم: ابتدا، فهرستی از مجراهای انتشار - شامل مجلات و کنفرانس‌ها - در حوزه‌های پژوهشی آرشیو، علم اطلاعات و علوم انسانی رقمی گردآوری کردیم. این فهرست به عنوان ضمیمه در دسترس است (جدول ۱). درحالی‌که اذعان داریم بسیاری از مجراهای انتشار و حوزه‌های دیگر. برای مثال، حقوق، پلیس و نظارت و پردازش

1. Machine Learning (ML)
2. Natural Language Processing (NLP)
3. Computer Vision

اسناد. نیز می‌توانند با دامنه گسترده این پژوهش مرتبط باشند، فهرست انتخابی خود را نقطه عزیمت مناسبی برای بازنمایی مباحث جاری درباره موضوع مورد نظر دانستیم. در این مجراهای انتشار، تمامی شماره‌های منتشرشده در شش سال گذشته، از ۲۰۱۵ تا ۲۰۲۰ با احتساب هر دو سال را تا جدیدترین شماره موجود بررسی کردیم.

انتخاب سال ۲۰۱۵ به عنوان نقطه آغاز، دو دلیل داشت. نخست، هوش مصنوعی نوین تا حد زیادی از پیشرفت‌های اخیر در حوزه موسوم به یادگیری عمیق و کاربرد شبکه‌های عصبی بهره می‌گیرد. اگرچه پژوهش درباره شبکه‌های عصبی به چندین دهه پیش بازمی‌گردد، این رویکرد زمانی به جریان اصلی پژوهش راه یافت که در سال ۲۰۱۲، سامانه‌ای مبتنی بر شبکه‌های عصبی کانولوشنی در «چالش بزرگ مقیاس بازشناسی دیداری ایمجنت»^۱، در مقایسه با نزدیک‌ترین رقیب مبتنی بر رویکرد غیرعصبی، بهبودی ۴۱ درصدی به دست آورد. با توجه به زمان لازم برای تأثیرگذاری این نوآوری‌ها بر حوزه‌ای نسبتاً دورتر، مانند نگهداری اسناد، آغاز مرور از سال ۲۰۱۵ منطقی به نظر می‌رسید. نگاهی گذرا به بازتاب عمومی هوش مصنوعی پیش از این تاریخ نیز برای تأیید تصمیم ما کافی بود. دوم، قصد داشتیم بر مباحث جاری و چشم‌اندازهای آینده تمرکز کنیم؛ از این رو، هوش مصنوعی نوین را مورد توجه قرار دادیم.

در فرایند بررسی، بر انتخاب آثاری تمرکز کردیم که به طور مستقیم یا به صراحت، به نظریه و عمل نگهداری اسناد اشاره داشتند؛ بنابراین، آثار صرفاً فنی درباره مواد آرشیوی را که فاقد دیدگاهی آرشیوی بودند و نیز آثار آرشیوی فاقد ارجاع صریح به هوش مصنوعی را وارد بررسی نکردیم. این فرایند به شناسایی مجموعه‌ای از آثار انجامید که آن‌ها را با جزئیات بیشتری بررسی و براساس مدل پیوستار اسناد سازمان دهی کردیم. علاوه بر این، این آثار را نقطه آغاز جست‌وجوی استنادی قرار دادیم؛ بدین ترتیب، فهرست منابع آن‌ها و استنادهای دریافت‌شده هر اثر را بررسی کردیم تا منابع مرتبط بیشتری را، حتی خارج از فهرست اولیه مجراهای انتشار، به مجموعه مورد بررسی بیفزاییم. در نهایت، این فهرست گسترده را با منابع کلیدی و جدیدی تکمیل کردیم که براساس تخصص خود مرتبط می‌دانستیم و خارج از ادبیات آرشیو و علم اطلاعات یا علوم انسانی رقمی قرار داشتند. حاصل این فرایند، گزینش ۵۳ منبع است که در این مقاله آن‌ها را تجزیه و تحلیل می‌کنیم. این پژوهش را نه یک مرور نظام‌مند، بلکه مروری کلی، نماینده و گویا از موضوع مورد بحث می‌دانیم و اذعان داریم که دیدگاه‌های خود ما نیز بر فرایند انتخاب و تحلیل منابع تأثیر گذاشته‌اند.

۴. بررسی آرشیوها و هوش مصنوعی

ادبیات را در چهار بخش موضوعی کلی سازمان دهی می‌کنیم: ملاحظات نظری و حرفه‌ای؛ فرایندهای نگهداری اسناد؛ سازمان دهی و دسترسی به آرشیوها؛ و شکل‌های جدید آرشیوهای رقمی. بخش اخیر، روندهای نوظهور را نیز دربر می‌گیرد. در هر بخش، آثار مرتبط را با استفاده از مدل پیوستار اسناد به عنوان چارچوب مرجع بررسی می‌کنیم.

یکی از پرسش‌های مهم این است که گسترش هوش مصنوعی چگونه بر مفاهیم بنیادین آرشیوی و نظریه آرشیوی تأثیر می‌گذارد. این بحث عمدتاً در بستر تحول کلی رقمی مطرح شده و کمتر به‌طور خاص از منظر هوش مصنوعی بررسی شده است. ماس^۱ (۲۰۱۸) تبدیل آرشیو از مجموعه‌ای از اسناد اداری به مجموعه‌ای از داده‌ها را یکی از بنیادی‌ترین پیامدهای داده‌سازی می‌داند. بازفهم آرشیو به‌مثابه داده‌های «معنادار»، ازجمله، به این معناست که روش‌های سنتی ارزشیابی دیگر کفایت نمی‌کنند. نه تنها به ابزارهای جدید مبتنی بر همکاری انسان و ماشین نیاز است، بلکه روش‌های تازه‌ای نیز برای استفاده و تجزیه و تحلیل آرشیوها به‌مثابه داده ضرورت دارد. آپوارد^۲ (۲۰۱۸) از «ورشکستگی محصولات مدیریت اسناد» سخن می‌گوید و چارچوب جدیدی با عنوان «انفورماتیک نگهداری اسناد» برای پشتیبانی از فرایندهای آرشیوی ارائه می‌کند. بازگرداندن «کتیبه‌های آبشاری»^۳ مستلزم آن است که متخصصان اطلاعات با «آرشیو نانوثانیه‌ای»^۴ که در آن همه چیز درهم تنیده است، درگیر شوند. این وضعیت، به تعبیر آپوارد، مستلزم «شکل‌های جدیدی از حرفه‌ای‌گری است که بتوانند با اقتدار با قدرت تخصیصی شکل‌های جدید محاسبه مواجه شوند» (Upward, 2018).

آگاهی گسترده‌ای وجود دارد مبنی بر اینکه تحول و دگرگونی رقمی، مفاهیم آرشیوی‌ای مانند منشأ و نظم اولیه را با چالش مواجه کرده است (Thibodeau, 2016). این تحول، بحث‌هایی را درباره ارتباط و معنای مفاهیم تثبیت‌شده آرشیوی، ازجمله «سند، شواهد، ثبات، ماندگاری، یگانگی، اصالت، مالکیت و حضانة» برانگیخته است (Gilliland, 2014). تیبودو^۵ استدلال می‌کند که این بحث‌ها اغلب به موضوعی غیرمولد می‌انجامند، درحالی‌که اتخاذ رویکردی متفاوت می‌تواند به غنای مفهومی این حوزه کمک کند. او نشان می‌دهد که چگونه می‌توان مفاهیم علم آرشیو را با ترکیب «زبان‌شناسی نقش‌گرایی نظام‌مند»^۶ - که چارچوبی برای فهم منشأ ازطریق تجزیه و تحلیل تجربی زمینه فراهم می‌کند. و نظریه گراف غنی کرد و بازنمایی نمادین تازه‌ای از آن‌ها ارائه داد. چنین رویکردی می‌تواند فهم اسناد و فرایندهای ایجاد و استفاده آن‌ها را بهبود بخشد.

بحث اساسی دیگر به نقش نگهداری اسناد در تأمین شفافیت و پاسخ‌گویی اسنادی مربوط می‌شود که بر مبنای خروجی‌های الگوریتمی شکل گرفته‌اند. این موضوع به ابعاد ایجاد و ثبت اسناد مربوط است. آندرسن^۷ (۲۰۲۰) چارچوبی برای بحث و یافتن شیوه‌های مستندسازی توضیحات مربوط به الگوریتم‌ها ارائه می‌دهد. این چارچوب بر مفاهیمی از مدیریت اسناد، فلسفه علم و مطالعات حقوقی استوار است. در همین راستا، بان^۸ (۲۰۲۰) موضوع را از منظر «هوش مصنوعی توضیح‌پذیر»^۹ بررسی می‌کند؛ رویکردی که نه تنها نگهداری اسناد، بلکه حفظ توضیحات مرتبط با آن‌ها را نیز ضروری می‌داند. شیرز^{۱۰} استدلال می‌کند که آرشیوها باید متناسب با این تحولات تغییر کنند؛ به‌ویژه آن‌که «مقررات عمومی حفاظت از داده‌ها»^{۱۱} خواستار ارائه «اطلاعات معنادار درباره منطق» دخیل در تصمیم‌گیری ماشینی است.

1. Moss
2. Upward
3. Cascading Inscriptions
4. Nanosecond Archive
5. Thibodeau
6. Systemic Functional Linguistics
7. Andresen
8. Bunn
9. Explainable Artificial Intelligence (XAI)
10. Shears
11. General Data Protection Regulation (GDPR)

با توجه به پیامدهای هوش مصنوعی برای این حرفه، تایمر^۱ (۲۰۱۸) تصویری از آینده حرفه‌ای ترسیم می‌کند که در آن ماشین‌ها و الگوریتم‌ها می‌توانند وظایف آرشیوی متعددی را بر عهده گیرند. این تحول بدین معناست که آرشیوداران باید به «استادان داده» تبدیل شوند، اما هم‌زمان، این حرفه باید بیش‌ازپیش بر «روایت، داستان‌گویی، معناسازی و ارائه زمینه» تمرکز کند؛ زیرا «اطلاع‌رسانی و آموزش، بیش‌از هر زمان دیگری اهمیت خواهند داشت» (Theimer, 2018). رولان^۲ (۲۰۱۹) به شکاف در حال ظهور میان آرشیودارانی اشاره می‌کند که با مفاهیم هوش مصنوعی آشنا هستند و کسانی که چنین آشنایی‌ای ندارند و بر ضرورت «بازآموزی» برای کاهش شکاف دانش و مهارت تأکید می‌کند.

لُمیو^۳ از شکل‌گیری یک «فرارشته» جدید حمایت می‌کند که آرشیوها را به حوزه محاسبات نزدیک کند و دانشمندان علوم رایانه و مهندسان نرم‌افزار را به حوزه آرشیو وارد سازد. این گفت‌وگو در حوزه نوظهوری بازتاب یافته است که با عنوان «علم آرشیو محاسباتی»^۴ شناخته می‌شود (Le-mieux, 2018; Marciano et al., 2018). راناده^۵ (۲۰۱۸) استدلال می‌کند که پیشرفت‌های فنی مستلزم تغییری چشمگیر در تفکر آرشیوی است؛ تغییری که قابلیت استفاده، اعتماد و زمینه را در مرکز فعالیت آرشیوی قرار می‌دهد. مؤلفه‌های اصلی این بازاندیشی عبارتند از پذیرش آشفتگی و عدم قطعیت و شفافیت درباره ماهیت احتمالاتی و عدم قطعیت موجود در داده‌ها و فراداده‌های آرشیوی. آپوارد (۲۰۱۸) نیز خواستار بنیانی رشته‌ای و جدید است که به صورت‌بندی‌های حرفه‌ای تازه و مهارت‌ها و دانش جدید نیاز دارد و درنهایت نتیجه می‌گیرد که «کلید تغییر در دست مربیان است» (Upward, 2018).

۴-۲. خودکارسازی فرایندها و تصمیم‌های نگهداری اسناد

یکی از چشم‌اندازهای امیدبخش فناوری‌های هوش مصنوعی، ظرفیت آن‌ها برای خودکارسازی بخش‌هایی از جریان کاری آرشیوی است. بحث‌ها و انتظارات مربوط به این ظرفیت عمدتاً در ابعاد «ثبت» و «تکثر» مدل پیوستار اسناد قرار می‌گیرند. اگرچه رولان و همکاران (۲۰۱۹) تغییرات عمیقی را در فعالیت‌های آرشیوی طی سال‌های آینده پیش‌بینی می‌کنند، پیشرفت‌های عملی در این زمینه را چندان رضایت‌بخش نمی‌دانند؛ وضعیتی که می‌توان آن را، ازجمله، به دشواری عبور از مرزهای رشته‌ای نسبت داد. طراحی کاربردهای دقیق هوش مصنوعی مستلزم داده، دانش تخصصی حوزه و همکاری دانشمندان داده‌ای است که الگوریتم‌ها را توسعه می‌دهند. علاوه بر این، نباید حجم قابل‌توجه کارهای مقدماتی - ازجمله آماده‌سازی داده‌ها و توسعه جریان‌های کاری - را که پیش‌از به‌کارگیری موفق هوش مصنوعی ضرورت دارند، دست‌کم گرفت؛ موضوعی که چهار مطالعه موردی در استرالیا نیز آن را نشان داده‌اند. در برخی موارد، مدیریت دستی، مشارکتی و باکیفیت داده‌ها نیز رویکردی موفق بوده است (Turchin, 2015).

1. Theimer
2. Rolan
3. Lemieux
4. Computational
Archival Science
5. Ranade

بخش عمده ادبیات بررسی شده درباره خودکارسازی فرایندهای آرشیوی بر مسائل ارزشیابی تمرکز دارد. اگرچه دستورالعمل‌های نوین نگهداری اسناد تأکید می‌کنند که ارزشیابی باید رویکردی پیش‌نگرانه، نه واکنشی، به ایجاد، ثبت و مدیریت اسناد داشته باشد، واقعیت آن است که متخصصان نگهداری اسناد با حجم روزافزونی از اسناد غیرساختارمند و طبقه‌بندی نشده مواجه‌اند. هوش مصنوعی از ظرفیت کمک به مدیریت چنین حجم عظیمی از اسناد برخوردار است و چندین پژوهشگر نیز نتایج آزمایش‌ها و پژوهش‌های خود را در این زمینه گزارش کرده‌اند.

ولینو^۱ (۲۰۱۶) آزمایش‌هایی را بر دو مجموعه ایمیل شرح می‌دهد و نشان می‌دهد که می‌توان طبقه‌بندهای خودکار را با موفقیت آموزش داد تا تصمیم‌های کارشناسان را در تشخیص ایمیل‌های دارای ارزش تجاری و فاقد این ارزش بازتولید کنند. لی^۲ (۲۰۱۸) پیشرفت‌هایی را در استفاده از پشتیبانی نرم‌افزاری در فرایندهای آرشیوی مشاهده می‌کند، اما استدلال می‌کند که بهره‌گیری از «پزشکی قانونی رقمی»^۳، پردازش زبان طبیعی و یادگیری ماشینی می‌تواند به توسعه پشتیبانی مؤثرتر در حوزه حیاتی ارزشیابی اسناد کمک کند؛ ازاین‌رو، خواستار پژوهش‌های بیشتر در این زمینه است.

هاچینسون^۴ (۲۰۲۰) با ارائه مروری کلی بر تلاش‌های صورت‌گرفته برای استفاده از پردازش زبان طبیعی و یادگیری ماشینی در ارزشیابی، این موضوع را بیشتر بررسی می‌کند. او چندین ابزار نرم‌افزاری، ازجمله ePADD، BitCurator NLP، ArchExtract و تعدادی ابزار تجاری را بررسی و پنج اصل طراحی-کاربردپذیری، تعامل‌پذیری، انعطاف‌پذیری، تکرارپذیری و قابلیت تنظیم-راشناسایی می‌کند که ابزارهای پردازش زبان طبیعی برای ادغام موفق در جریان‌های کاری آرشیوی باید از آن‌ها پیروی کنند. شابو^۵ و همکاران (۲۰۲۰) پژوهشی اکتشافی را گزارش می‌کنند که در آن، اثبات مفهوم و نمونه اولیه ابزاری برای ارزشیابی آرشیوی توسعه یافت. هدف این ابزار، پشتیبانی از شناسایی و استخراج داده‌های ساختاریافته و غیرساختاریافته سازمانی به‌منظور امحای به‌موقع یا انتقال آن‌ها به مخزن آرشیوی بود. پژوهشگران ابزاری طراحی کردند که رویکرد آرشیوی از بالا به پایین - شامل چارچوبی از ابعاد، متغیرها و معیارها - را با رویکرد داده‌کاوی از پایین به بالا - شامل طبقه‌بندی خودکار موضوعات، نمایه‌سازی کامل و بازشناسی موجودیت‌های نامدار - ترکیب می‌کرد تا ارزشیابی حجم زیادی از اسناد و داده‌های رقمی را تسهیل کند.

۴-۴. رسیدگی به اطلاعات حساس

یکی از فوری‌ترین چالش‌های پیش روی سازمان‌های دولتی و مؤسسات آرشیوی، جلوگیری از افشای زودهنگام اطلاعات حساس یا اطلاعات قابل‌شناسایی شخصی است. درعین‌حال، سازمان‌ها می‌کوشند از وضعیتی جلوگیری کنند که در آن، ناتوانی در شناسایی اطلاعات حساس، پاسخ‌گویی به «درخواست‌های آزادی اطلاعات»^۶ یا فراهم‌کردن امکان پژوهش آرشیوی را عملاً

1. Vellino
2. Lee
3. Digital Forensics
4. Hutchinson
5. Shabou
6. Freedom of Information (FOI)

ناممکن سازد. بارون^۱ و پین^۲ (۲۰۱۷) رویکردها و فنون تحلیلی گوناگونی را برای شناسایی محتوای حساس پیشنهاد می‌کنند و دستورکاری پژوهشی برای جامعه آرشیوی، با هدف توسعه راهبردهای مبتنی بر پشتیبانی فناوری، ارائه می‌دهند.

اشنایدر^۳ و همکاران (۲۰۱۹) توضیح می‌دهند که پنج مؤسسه، چگونه از نرم‌افزار ePADD برای غربالگری ایمیل‌ها و شناسایی اطلاعات حساس، محرمانه یا مشمول محدودیت‌های قانونی استفاده می‌کنند. به همین ترتیب، ماس و همکاران (۲۰۱۷) از توسعه ابزارهای کمکی برای بازیابی حساسیت در محیط رقمی حمایت می‌کنند، اما هم‌زمان، تلقی ساده‌انگارانه از مفهوم حساسیت را به چالش می‌کشند. از منظر آنان، نمی‌توان صرفاً موضوع یک سند را معیار تعیین حساس بودن آن دانست؛ زیرا حساسیت تا حد زیادی به زمینه‌هایی وابسته است که اسناد در آن‌ها ایجاد و استفاده می‌شوند. از این رو، نویسندگان خواستار پژوهش‌هایی هستند که تحلیل زمینه‌ای را دربرگیرند تا تصمیم‌گیری‌های دقیق‌تر و ظریف‌تر درباره حساسیت اسناد امکان‌پذیر شود.

به همین ترتیب، سوزا^۴ و همکاران (۲۰۱۶) دشواری‌های شناسایی اطلاعات حساس را بررسی و رویکردی برای طبقه‌بندی خودکار اسناد در آرشیوهای دولتی ارائه می‌کنند تا مشخص شود آیا اسناد قابلیت انتشار دارند یا باید همچنان طبقه‌بندی شده باقی بمانند. هاجینسون (۲۰۱۸) پیشنهاد می‌کند که «یادگیری ماشینی نظارت‌شده»^۵ می‌تواند رویکرد مناسبی برای «تریاج»^۶ مجموعه‌ها و شناسایی اسناد حساس از منظر حریم خصوصی باشد.

۴-۵. فراداده‌ها

یکی دیگر از حوزه‌های فرایندهای نگهداری اسناد که کاربرد هوش مصنوعی در آن مورد بحث قرار می‌گیرد، توصیف و استخراج فراداده‌ها است. بوتنر^۷ و همکاران (۲۰۱۹) از پروژه‌ای کوچک، اما موفق، برای اثبات مفهوم گزارش می‌دهند که در آن سامانه‌ای برای طبقه‌بندی خودکار مبتنی بر قواعد معنایی اسناد رسمی شورای اروپا توسعه یافت. هدف این سامانه، کاهش ضرورت ورود دستی فراداده‌ها از سوی تولیدکنندگان اسناد و ارتقای کیفیت فراداده‌های موضوعی بود. اسپنسر^۸ (۲۰۱۷) نیز توضیح می‌دهد که چگونه می‌توان از روش‌های خودکار برای شناسایی روابط میان اسناد استفاده کرد و این فناوری‌ها چگونه می‌توانند به فراهم کردن زمینه بیشتر برای فهم اسناد کمک کنند.

۵. سازمان‌دهی و دسترسی به آرشیوها

افزایش روزافزون مواد آرشیوی رقمی و زاده‌شده رقمی، خودکارسازی و حتی بازاندیشی اساسی در نحوه سازمان‌دهی اسناد و دسترسی به آن‌ها را ضروری ساخته است. چندین اثر مرتبط با این موضوع در ابعاد «سازمان‌دهی» و «تکثر» مدل پیوستار اسناد قرار می‌گیرند؛ بدین معنا که وجود اسنادی را مفروض می‌گیرند که پیش‌تر ایجاد و ثبت شده‌اند. با این حال، این فرایندها اغلب خود به ایجاد اطلاعات آرشیوی جدید می‌انجامند.

1. Baron
2. Payne
3. Schneider
4. Sousa
5. Supervised Machine Learning
6. Triage
7. Büttner
8. Spencer

یکی از مضامین مشترک این است که می‌توان از هوش مصنوعی برای فراتر رفتن از اصول سنتی سازمان‌دهی و دسترسی به آرشیوها، مبتنی بر اصل منشأ و نظم اولیه، بهره‌گرفت؛ اصولی که عمدتاً بر فهرست‌ها و توصیف واحدهای آرشیوی تمرکز دارند. امروزه می‌توان این رویکرد را با ابزارهای دیگری تکمیل کرد که عمدتاً بر محتوای اسناد متمرکزند. این تحول نه تنها به کاربران امکان می‌دهد حجم گسترده‌تری از مواد آرشیوی را درک و بررسی کنند، بلکه شیوه‌های نوینی را نیز برای مواجهه با این مواد در اختیار آنان قرار می‌دهد.

۵-۱. استخراج و نمایه‌سازی خودکار محتوا

بازاندیشی در نحوه سازمان‌دهی آرشیوها با استفاده از هوش مصنوعی، از استخراج خودکار محتوا و نمایه‌سازی آن آغاز می‌شود. اگرچه در اینجا تمرکز ما بر پردازش فنی اسناد نیست. برای مروری جدید در این زمینه بنگرید به بینماخاشن^۱ و محمود^۲ (۲۰۲۰). نقش Transkribus را دقیقاً در همین زمینه برجسته می‌کنیم (Transkribus, Muehlberger et al., 2019). ابزاری است که مدل‌های بازشناسی تصویر و متن را برای تسهیل «بازشناسی متن دست‌نویس»^۳ و «بازشناسی نوری نویسه»^۴ ادغام می‌کند. این ابزار که از طریق رابط گرافیکی یا به صورت برنامه‌نویسی و با استفاده از «رابط‌های برنامه‌نویسی کاربردی»^۵ در دسترس است، به گونه‌ای طراحی شده است که در جریان‌های کاری آرشیوی ادغام شود و از مخازن روبه‌رشد تصاویر رقمی متون تاریخی بهره‌گیرد. Transkribus به‌طور فزاینده‌ای در آرشیوها و دیگر مؤسسات به کار گرفته می‌شود و در گسترش کاربرد هوش مصنوعی برای استخراج محتوای اسناد و جست‌وجویکردن آن‌ها نقش مهمی ایفا می‌کند. دسترسی به محتوای اسناد، به‌ویژه متون، امکان استفاده از آن‌ها برای نمایه‌سازی را فراهم می‌کند. کولاویتزا^۶ (۲۰۱۹) پیشنهاد می‌کند که سامانه‌های اطلاعات آرشیوی با استفاده از نمایه‌های تاریخی که غالباً در دوره فعال بودن آرشیو تولید شده‌اند، به‌طور خودکار ایجاد شوند. از آنجا که این نمایه‌ها اغلب بر موجودیت‌هایی مانند اشخاص، مکان‌ها یا کلیدواژه‌هایی تمرکز دارند که در سراسر مجموعه‌های آرشیوی ظاهر می‌شوند، به کاربران امکان می‌دهند آرشیو را به شیوه‌ای مکمل اصل منشأ و نظم اولیه جست‌وجو کنند. هدفی مشابه، نقطه عزیمت پروژه Traces through Time در آرشیو ملی بریتانیا است که بر اشخاص به عنوان موجودیت تمرکز دارد (Ranade et al., 2016). نتایج این پروژه نشان داد که نمایه‌سازی مبتنی بر موجودیت، با ایجاد پیوند میان مجموعه‌های آرشیوی مختلف، امکان جست‌وجوی «روان‌تر» در آرشیو را فراهم می‌کند. نمایه‌سازی مبتنی بر موجودیت نیز اغلب در قالب «داده‌های پیوندی»^۷ ارائه می‌شود (de Wilde & Hengchen, 2017).

۵-۲. خوانش دور اسناد آرشیوی

جریان روبه‌رشدی از پژوهش‌ها بر «خوانش دور»^۸ محتوای اسناد تمرکز دارد (Moretti, 2013). برای نقد این مفهوم و رویکردهای مرتبط با آن در زمینه مطالعات ادبی نیز بنگرید به بود^۹ (۲۰۱۷). یکی از روش‌های رایج در این زمینه، «مدل‌سازی موضوعی»^{۱۰} است که به کاربران امکان می‌دهد

1. Binmakhashen
2. Mahmoud
3. Handwritten Text Recognition (HTR)
4. Optical Character Recognition (OCR)
5. Application Programming Interfaces (APIs)
6. Colavizza
7. Linked Data
8. Distant Reading
9. Bode
10. Topic Modeling

«موضوعات» موجود در مجموعه‌های بزرگ متنی را شناسایی کنند؛ موضوعاتی که براساس هم‌وقوعی مکرر اصطلاحات تعریف می‌شوند.

چنی^۱ (۲۰۱۶) مدل «کپسول» را برای تجزیه و تحلیل پیکره‌های تاریخی حاوی رویدادها ارائه می‌کند و تلگراف‌های دیپلماتیک وزارت امور خارجه ایالات متحده را به عنوان مطالعه موردی بررسی می‌کند. روش پیشنهادی او امکان شناسایی رویدادهای استثنایی‌ای را فراهم می‌کند که از زبان متعارف دیپلماتیک فاصله می‌گیرند. به همین ترتیب، بلانکه^۲ و همکاران (۲۰۱۷) از مدل‌سازی موضوعی و مدل‌های زبانی، به‌ویژه «جاسازی واژه‌ها»^۳، برای استخراج و مطالعه «دوره‌ها» در آرشیوهای متنی استفاده می‌کنند. آنان «دوره» را بازه‌ای از کاربرد منسجم زبان تعریف می‌کنند. دوره‌ها ابتدا با استفاده از معیارهای مبتنی بر واژه‌نامه شناسایی و سپس با بهره‌گیری از مدل‌های موضوعی و جاسازی واژه‌ها به صورت کیفی تحلیل می‌شوند.

کاربرد دیگر مدل‌های موضوعی، استخراج کلیدواژه‌ها برای نمایه‌سازی است. هنگشن^۴ (۲۰۱۶) این روش را در آرشیو کمیسیون اروپا به کار می‌گیرد. در این رویکرد، کلیدواژه‌ها استخراج و سپس با اصطلاح‌نامه موجود تطبیق داده می‌شوند تا قابلیت کشف این منابع افزایش یابد. کلیدواژه‌های استخراج شده را می‌توان با استفاده از مدل‌های زبانی نیز غنی کرد تا نمایه‌سازی معنایی محتوای آرشیوی بیش‌ازپیش تقویت شود (Coeckelbergs et al., 2020).

با توجه به ماهیت عمدتاً متنی اسناد آرشیوی، خوانش دور، جایگاه برجسته‌ای دارد؛ باین‌حال، پژوهش درباره دیگر شکل‌های محتوا نیز درحال گسترش است. برای مثال، فان دن هوفل^۵ و همکاران (۲۰۱۵) به‌طور مستدل از نمایه‌سازی هم‌زمان متون و محتوای بصری، مانند نقاشی‌هایی که ازطریق شبکه‌های مکانباتی در سراسر «جمهوری ادبیات» گردش می‌کردند، دفاع می‌کنند. وورز^۶ و اسمیت^۷ (۲۰۱۹) نیز مروری روشن و در دسترس بر ظرفیت‌های هوش مصنوعی نوین برای «مشاهده دور»^۸ ارائه می‌دهند.

۳-۵. جست‌وجو و بازیابی

جست‌وجو و بازیابی اطلاعات موجود در اسناد آرشیوی نیز حوزه فعالی در پژوهش بوده است. لی و همکاران (۲۰۱۹) از تطبیق الگوی بصری و طبقه‌بندی خودکار برای بازیابی کارت‌های مرجع از آرشیو رقمی «خدمت بین‌المللی ردیابی»^۹ استفاده می‌کنند؛ آرشیوی که یکی از بزرگ‌ترین و ناهمگون‌ترین مجموعه‌های مرتبط با هولوکاست را دربر دارد. ضرورت کمک به آرشیوداران و دیگر کاربران برای غربال اسناد بالقوه مرتبط ازطریق طبقه‌بندی خودکار، در چارچوب‌های پیش‌بینی دیگری نیز مورد توجه قرار گرفته است (Risi et al., 2019).

پل^{۱۰} و همکاران (۲۰۲۰) درمقابل، رویکرد متفاوتی را مبتنی بر فهرست‌های آرشیوی اتخاذ می‌کنند. آنان پیشنهاد می‌کنند که با استفاده از فنون متن‌کاوی، توصیف‌های آرشیوی به‌طور خودکار تجمیع شوند تا کاوش و کشف آن‌ها تسهیل شود. پروژه EHR^{۱۱} نیز از روشی مکمل بهره

1. Cheney
2. Blanke
3. Word Embeddings
4. Hengchen
5. Van den Heuvel
6. Wevers
7. Smits
8. Distant Viewing
9. International Tracing Service (ITS)
10. Bell
11. European Holocaust Research Infrastructure (EHRI)

می‌گیرد. این پروژه، فراداده‌های آرشیوی چندین منبع را به‌طور خودکار همگام‌سازی و آن‌ها را با فراداده‌های تولیدشده ازسوی کاربران و «مجموعه‌های مجازی» موضوعی تکمیل می‌کند.

فراداده‌های آرشیوی همچنین برای سازمان‌دهی مواد آرشیوی به شیوه‌هایی نوین‌تر و غنی‌تر به کار می‌روند. یکی از نمونه‌ها، استفاده از «تحلیل احساس»^۱ در آرشیو ژرمن گریر^۲ است (Weber, 2019). بوسیت^۳ و همکاران (۲۰۲۰) نیز در چارچوب پروژه Re:TV، بازایی شخصی‌سازی‌شده را به‌عنوان یکی از کاربردهای هوش مصنوعی برای پشتیبانی از تعامل کاربران با مجموعه‌های صوتی و تصویری بررسی می‌کنند. علاوه‌بر این، روش‌های هوش مصنوعی بیش‌ازپیش در ابزارهای مورد استفاده مورخان ادغام می‌شوند و به کاربران امکان می‌دهند محتوای آرشیوی را از طریق «تجسم اطلاعات»^۴ کاوش کنند (Moretti et al., 2016; Lansdall-Welfare et al., 2020).

کاربرد هوش مصنوعی در جست‌وجو و بازایی ازسوی مورخان نیز به‌طور انتقادی بررسی شده است. وینترز^۵ (۲۰۱۹) با مسئله‌مندکردن استفاده از جست‌وجوی تمام‌متن در آرشیوهای زاده‌شده رقمی، خواستار رویکردهای جدیدی در علم آرشیو و هوش مصنوعی است که حساسیت بیشتری به سلسله‌مراتب و زمینه آرشیوی داشته باشند. چنین رویکردهایی باید از افتادن در دام یافتن صرف آنچه از پیش شناخته شده است. یا «سوگیری تأیید». اجتناب کنند و در عوض، جست‌وجوی ناشناخته‌ها و شناسایی شکاف‌ها و غیبت‌های موجود در داده‌ها را مورد توجه قرار دهند. ازاین‌منظر، این دیدگاه با دیگر تأملات نوظهور درباره آرشیوهای زاده‌شده رقمی (Chabin, 2020) و کاربرد تفکر آرشیوی در یادگیری ماشینی (Jo & Gebru, 2019) که در ادامه به آن‌ها خواهیم پرداخت، ارتباط می‌یابد.

یکی از راه‌حل‌های ممکن، «جست‌وجوی انتقادی»^۶ است (Guldi, 2018). گودی^۷ (۲۰۱۸) پیشنهاد می‌کند که به‌جای تمرکز بر کاربرد یک ابزار واحد برای بازایی اطلاعات از آرشیوهای تاریخی، نوعی «نقد ابزار» در هر مرحله فرایند جست‌وجوی مورخ در آرشیوهای رقمی به کار گرفته شود. این فرایند از مجموعه‌ای از مراحل متناوب کاربرد الگوریتم‌ها، خواندن و تأمل انتقادی تشکیل می‌شود.

۶. اشکال نوین آرشیوهای رقمی

گسترش فناوری‌های رقمی در تمامی عرصه‌های جامعه، علاوه‌بر فراهم‌کردن فرصت‌های جدید برای خودکارسازی ابعاد مختلف فرایندهای آرشیوی، به شکل‌گیری اشکال نوینی از آرشیوهای رقمی نیز انجامیده است. این تحول، حوزه «ثبت» و «سازمان‌دهی» آرشیوها به‌مثابه نهاد را به ابعاد «ایجاد» و «تکثر» مدل پیوستار اسناد گسترش می‌دهد. فناوری‌های هوش مصنوعی در این فرایند نقش‌های متعددی ایفا می‌کنند. نخست، این فناوری‌ها شکل‌های جدیدی از گردآوری داده را تسهیل می‌کنند؛ ازجمله سامانه‌های تعبیه‌شده و نیمه‌هوشمند «اینترنت اشیا»^۸. سودرینگ^۹ (۲۰۲۰) براساس دو مورد کاربردی - داده‌های حسگر کیفیت آب و بازشناسی خودکار شماره پلاک - استدلال می‌کند که سازمان‌های دولتی استفاده‌کننده از چنین سامانه‌های اینترنت اشیا باید

1. Sentiment Analysis
2. Germaine Greer
3. Bocyte
4. Information Visualization
5. Winters
6. Critical Search
7. Guldi
8. Internet of Things (IoT)
9. Södring

اطمینان یابند که نحوه ثبت و سازمان دهی اطلاعات حاصل با قوانین، مقررات و انتظارات مرتبط با حریم خصوصی و دیگر ارزش های عمومی سازگار است. نویسندگان بر این باورند که هشت اصل پذیرفته شده نگهداری اسناد. پاسخ گویی، شفافیت، یکپارچگی، حفاظت، انطباق، دسترسی پذیری، نگهداری و تعیین تکلیف. می توانند رهنمودهایی برای سازمان دهی این اشکال جدید آرشیوهای رقمی فراهم کنند.

به همین ترتیب، شابین^۱ (۲۰۲۰) نشان می دهد که چگونه اصول علم آرشیو می توانند به ارتقای کیفیت گردآوری و پردازش داده ها در خارج از زمینه های رسمی و تعریف شده نگهداری اسناد کمک کنند. او با بررسی مجموعه اسنادی که دولت فرانسه در جریان «بحث بزرگ ملی»^۲ - برنامه راه اندازی شده در واکنش به اعتراض های جنبش موسوم به «جلیقه زردها»^۳ در اوایل سال ۲۰۱۹. گردآوری کرد، نشان می دهد که اتخاذ رویکردی «سندشناختی»^۴ برای احراز اصالت اسناد چگونه می تواند استخراج دانش درباره جمعیت مشارکت کنندگان و رفتار آنان را امکان پذیر سازد. علاوه بر این، شابین استدلال می کند که افزودن «رویکردی بدون نظارت»^۵ به تجزیه و تحلیل این آرشیو زاده شده رقمی می تواند «تنوع و غنای صدای شهروندان را در سراسر آن» آشکار سازد (Chabin, 2020)؛ از جمله موضوعات و پیشنهادهایی که در چارچوب مرجع پرسش های رسمی دولت که در آغاز مبنای تحلیل قرار داشتند، گنجانده نشده بودند. در اینجا، فناوری ها و روش های هوش مصنوعی به عنوان ابزاری برای تحقق شکل دموکراتیک تر و فراگیرتری از گردآوری و تجزیه و تحلیل داده ها مطرح می شوند.

فناوری های هوش مصنوعی همچنین امکان ثبت و سازمان دهی اسناد مربوط به رویدادهایی را فراهم می کنند که معمولاً در محدوده سیاست های قانونی یا خودتعریف شده گردآوری آرشیوهای سازمانی قرار نمی گیرند. برای مثال، سونمز^۶ (۲۰۱۶) سامانه ای را برای خزش خودکار منابع رقمی مرتبط با اعتراض های اجتماعی در برزیل توصیف می کند. در این سامانه، یک طبقه بند، گزارش های مرتبط با اعتراض ها را در «خوراک های خبری»^۷ شناسایی می کند و از «استخراج موجودیت»^۸ برای نمایه سازی اسناد بهره گرفته می شود. حاصل این فرایند، پایگاه داده جدیدی از اعتراض هاست که رویدادهایی مانند اعتصاب، تظاهرات، تحریم، تجمع های اعتراضی و شورش و نیز ویژگی های آن ها، از جمله مکان، مشارکت کنندگان و ایدئولوژی را دربر می گیرد.

در موارد دیگر، فناوری ها و روش های هوش مصنوعی برای تجزیه و تحلیل مجموعه های جدید و گسترده اسناد رقمی به کار گرفته می شوند که در داخل و خارج از مؤسسات آرشیوی تثبیت شده شکل گرفته اند. کانلی^۹ (۲۰۲۰) کاربرد فنون پردازش زبان طبیعی برای استخراج فراداده از اسناد دیپلماتیک موجود در پایگاه داده Freedom of Information Archive را بررسی می کند. این پژوهش ظرفیت رویکرد مذکور را با ایجاد شاخصی جدید برای سنجش «اهمیت کشور» در چارچوب اولویت های سیاست خارجی ایالات متحده نشان می دهد.

میلیگان^{۱۰} (۲۰۱۶) آرشیوهای وب، مانند Internet Archive را برخوردار از ظرفیت چشمگیری

1. Chabin
2. Grand Débat National
3. Gilets Jaunes
4. Diplomats
5. Unsupervised Approach
6. Sönmez
7. News Feeds
8. Entity Extraction
9. Connelly
10. Milligan

برای پژوهش درباره تاریخ معاصر می‌داند؛ زیرا اگرچه وبگاه‌ها نماینده تمامی جمعیت نیستند، صدای میلیون‌ها فرد غیرنخبه را ثبت می‌کنند. بلک^۱ (۲۰۱۶) نیز کاربرد «گردآوری ساختاریافته خودکار از وب»^۲ را به عنوان ابزاری برای تجزیه و تحلیل داده‌های آرشیوی وب بررسی می‌کند؛ رویکردی که به پژوهشگران امکان می‌دهد دیدگاه علوم انسانی را در مطالعه اینترنت به کار گیرند. او همچنین استدلال می‌کند که تجربه‌های حاصل از گردآوری ساختاریافته خودکار از وب می‌توانند برای متن‌کاوی مجموعه‌های بزرگ آرشیوی رقمی، مانند مجموعه‌های روزنامه‌ها، مفید باشند.

پروژه Archives Unleashed با فراهم کردن دسترسی به مجموعه‌های موضوعی وبگاه‌هایی که با ابزار Archive-It متعلق به Internet Archive آرشیو شده‌اند و نیز ارائه مجموعه‌ای از ابزارها برای استفاده از این منابع در جریان‌های کاری پژوهشی، در پی «در دسترس قرار دادن پتابیت‌ها محتوای تاریخی اینترنت برای پژوهشگران و دیگر علاقه‌مندان به مطالعه گذشته نزدیک» است (Ru-est et al., 2020). لینچ^۳ (۲۰۱۷) توجه را از داده‌ها و رویدادها به ابزارهای مورد استفاده برای ثبت و پردازش آن‌ها معطوف می‌کند. او مسئله «حفاظت و سرپرستی الگوریتم‌ها»^۴ را مطرح و استدلال می‌کند که فنون آرشیوی و آموزش‌های کنونی برای حفظ تصویری از «عصر الگوریتم‌ها» برای آینده مناسب نیستند. لینچ همچنین پیشنهادها و اولیه‌ای درباره چگونگی تحول آموزش آرشیوی ارائه می‌دهد تا آرشیوها بتوانند آنچه الگوریتم‌ها با ما انجام می‌دهند، ثبت و مستند کنند.

۶-۱. روندهای نوظهور

توجه به نقش هوش مصنوعی در فرایندهای آرشیوی به جامعه نگهداری اسناد محدود نمی‌شود. در سال‌های اخیر، این موضوع از منظر توسعه و پیاده‌سازی فناوری‌های هوش مصنوعی نیز مورد توجه قرار گرفته است؛ به ویژه از سوی پژوهشگرانی که اخلاق هوش مصنوعی را به مثابه یک «عمل اجتماعی-فنی»^۵ بررسی می‌کنند.

گوپتا^۶ و همکاران (۲۰۲۰) پیامدهای نظام‌های آرشیوی متمرکز و عادلانه را برای جوامع به حاشیه رانده شده. شامل افرادی که براساس فرهنگ، نژاد، جنسیت، سن یا توانایی‌های جسمی و ذهنی به حاشیه رانده شده‌اند. بررسی می‌کنند. آنان پیشنهاد می‌کنند نظام‌های هوش مصنوعی به گونه‌ای توسعه یابند که بتوانند فضاهای کم‌ترافیک و پراکنده اینترنت را شناسایی کنند؛ فضاهایی که برخی جوامع کوچک‌تر در آن‌ها آرشیوهای خود را ایجاد و نگهداری کرده‌اند. این دیدگاه نیز به ظرفیت استفاده از هوش مصنوعی برای شکل دادن به آرشیوهایی عادلانه‌تر و فراگیرتر اشاره دارد.

با این حال، جو^۷ و گبرو^۸ (۲۰۱۹) به استانداردهای انتخاب و توصیف آرشیوی توجه می‌کنند که طی قرن‌ها فعالیت در حوزه نگهداری اسناد توسعه یافته‌اند و ظرفیت آن‌ها را به عنوان الگویی برای ایجاد داده‌های آموزشی اجتماعی. فرهنگی کم‌سوگیرانه‌تر، شفاف‌تر و فراگیرتر در یادگیری ماشینی بررسی می‌کنند. در نهایت، محمد^۹ و همکاران (۲۰۲۰) دهه‌ها تفکر انتقادی در علوم انسانی را

1. Black
2. Automated Structured Web Harvesting
3. Lynch
4. Algorithm Curation
5. Socio-Technical Practice
6. Gupta
7. Jo
8. Gebru
9. Mohamed

به‌طور کلی بررسی و استدلال می‌کنند که این دانش می‌تواند چارچوبی مفهومی برای توسعه و پیاده‌سازی هوش مصنوعی فراگیرتر و خودمختارتر از منظر فرهنگی فراهم کند.

۷. بحث

مرور ما چندین روند قابل توجه را آشکار کرد. به گفته ماس (۲۰۱۸) تحول رقمی درحال بازپیکربندی آرشیو و تبدیل آن از مجموعه‌ای از اسناد اداری به مجموعه‌ای از داده‌هاست. درواقع، اصول و مفاهیم سنتی آرشیوی تحت فشار قرار گرفته‌اند و هوش مصنوعی اغلب برای آرمودن مرزهای آن‌ها و حتی فراتر رفتن از این مرزها به کار گرفته می‌شود. نمونه‌های برجسته این وضعیت، اصل منشأ و نظم اولیه‌اند که به‌طور سنتی برای سازمان‌دهی اسناد و دسترسی به آن‌ها استفاده می‌شوند. با این حال، با توجه به آرشیوهای عظیم رقمی شده یا زاده‌شده رقمی که امروزه با آن‌ها مواجه‌ایم، این اصول تنها یکی از رویکردهای ممکن برای سازمان‌دهی و دسترسی به اسناد به شمار می‌روند. از این منظر، هوش مصنوعی برای ایجاد نظم‌بندی‌های جایگزین و انعطاف‌پذیرتر در آرشیو به کار گرفته می‌شود؛ برای مثال، از طریق نمایه‌سازی آرشیو براساس محتوای اسناد. درحالی‌که محدودیت‌های اصول آرشیوی آشکارتر می‌شوند، همین اصول می‌توانند در جهان داده‌مند شده نیز کارکردهای تازه‌ای بیابند. چالش‌های نوظهور در جامعه هوش مصنوعی، ضرورت به کارگیری دانش و تجربه نگهداری اسناد درباره داده‌هایی را مطرح کرده‌اند که سامانه‌های نوین هوش مصنوعی بر آن‌ها متکی‌اند (Jo & Gebru, 2019). منشأ داده‌ها، ارزشیابی، زمینه‌مندسازی، شفافیت و پاسخ‌گویی از جمله موضوع‌هایی‌اند که آرشیوداران درباره آن‌ها از سنتی غنی و دانشی ارزشمند برخوردارند. پاسخ‌گویی به این ضرورت‌ها مستلزم بازاندیشی در آموزش حرفه‌ای آرشیوداران است. مرور ما نشان می‌دهد که اگرچه این نیاز به رسمیت شناخته شده، هنوز راه‌حل‌های مناسبی برای آن ارائه نشده است.

با توجه به کاربرد هوش مصنوعی در فرایندهای آرشیوی، با حوزه فعالیت‌های غنی و روبه‌رشد مواجه‌ایم که بر خودکارسازی این فرایندها تمرکز دارد. هوش مصنوعی برای خودکارسازی جزئی یا کامل فرایندهای کلیدی آرشیوی مانند ارزشیابی و ایجاد فراداده، آزمایش می‌شود. اگرچه این تلاش‌ها تاکنون نتایج عملی به همراه داشته‌اند یا ظرفیت دستیابی به چنین نتایجی را نشان می‌دهند، همچنان رویکردی بازاندیشانه و انتقادی به هوش مصنوعی کمتر مشاهده می‌شود. محدودیت‌های هوش مصنوعی در خودکارسازی فرایندهای آرشیوی و نیز پیامدهای گاه ناخواسته آن، هنوز به‌طور گسترده در ادبیات بررسی و واکاوی نشده‌اند. یکی از نمونه‌های مهم، سوگیری‌هایی است که ممکن است از تصمیم‌های مبتنی بر هوش مصنوعی ناشی شوند و پیامدهای اخلاقی به همراه داشته باشند. گام‌های اولیه امیدوارکننده در این زمینه (e.g., Bunn, 2020) نباید ما را از این واقعیت غافل سازند که بخش عمده بحث درباره پیامدهای فنی، اخلاقی و اجتماعی کاربرد هوش مصنوعی در فرایندهای آرشیوی هنوز شکل نگرفته است.

ملاحظه مشابهی درباره کاربرد هوش مصنوعی برای سازمان‌دهی آرشیوها و دسترسی به آن‌ها به شیوه‌های نوین مطرح است. مجموعه‌ای غنی از پژوهش‌ها، زمینه بازاندیشی در چگونگی تجربه آرشیوها، استفاده از آن‌ها و بهره‌گیری از اطلاعات موجود در آن‌ها را فراهم کرده است. اگرچه ادغام مؤثرتر فنون هوش مصنوعی در سامانه‌های اطلاعات آرشیوی همچنان نیازمند توسعه است، آثار بررسی‌شده در این مرور نشان می‌دهند که هوش مصنوعی می‌تواند برای پاسخ‌گویی بهتر به نیازهای طیف متنوع‌تری از کاربران به کار گرفته شود. یکی از مسیرهای آینده پژوهش که به نظر ما هنوز کمتر بررسی شده است، پیامدهای کاربرد هوش مصنوعی در سازمان‌دهی آرشیوها و دسترسی پژوهشگران و محققان دانشگاهی به آن‌هاست. به‌ویژه، ظرفیت گسترده‌ای برای طراحی و توسعه راه‌حل‌های مبتنی بر هوش مصنوعی وجود دارد که بتوانند نحوه استفاده پژوهشگران از آرشیوها را بهبود بخشند و غنی‌تر سازند. تحقق این هدف مستلزم گسترش تعامل‌های میان‌رشته‌ای میان آرشیوداران، پژوهشگران هوش مصنوعی و پژوهشگران علوم انسانی، مانند مورخان است (see Gul-di, 2018; Winters, 2019).

تحولات رقمی و به‌ویژه، وب، مفهوم آنچه سند محسوب می‌شود یا می‌تواند سند تلقی شود را نیز گسترش داده‌اند. اشکال نوین آرشیوهای زاده‌شده رقمی، از اینترنت اشیا گرفته تا رسانه‌های اجتماعی، نه تنها برای سازمان‌دهی و دسترسی، بلکه حتی برای ایجاد و ثبت اولیه به هوش مصنوعی متکی‌اند (e.g., Ruest et al., 2020). هوش مصنوعی همچنین ظرفیت خود را برای پشتیبانی از شکل‌های دموکراتیک‌تر و فراگیرتر آرشیوسازی نشان داده است؛ برای مثال، از طریق آشکارسازی، حفظ و دسترس‌پذیرکردن اسناد متعلق به جوامعی که بازنمایی کمتری یافته‌اند و گروه‌های اقلیت (e.g., Gupta & Kapoor, 2020). این اشکال نوین آرشیو، هم‌اکنون ضرورت بازاندیشی در فرایندهای آرشیوی و هم‌اکنون به مقیاس عظیم آن‌ها، چالش‌ها و فرصت‌هایی بی‌سابقه برای آرشیوداران ایجاد می‌کنند.

با استفاده از مدل پیوستار اسناد به عنوان چارچوب تحلیلی، دریافتیم که بخش بزرگی از آثار در دامنه این مرور بر ابعاد «سازمان‌دهی» و «تکثر» تمرکز دارند؛ در مرتبه بعد، بُعد «ثبت» قرار می‌گیرد و کمترین توجه به بُعد «ایجاد» معطوف شده است. این وضعیت تاحدی قابل درک است؛ زیرا هوش مصنوعی در درجه نخست برای خودکارسازی فرایندها در آرشیوهای موجود - برای مثال، پس از رقمی‌سازی یا در آرشیوهای زاده‌شده رقمی - و بهبود دسترسی به آن‌ها به کار گرفته می‌شود. با این حال، براساس مدل پیوستار اسناد، اسناد همواره در «وضعیت پیوسته شدن» قرار دارند. حتی هنگامی که هوش مصنوعی در بیرونی‌ترین ابعاد این مدل به کار گرفته می‌شود، آرشیوداران نیز به‌نوبه خود اطلاعات آرشیوی جدیدی ایجاد می‌کنند. به رسمیت شناختن این «پیوستار» و بیش از آن، ادغام این منطق در فرایندها و سامانه‌های آرشیوی، هنوز تاحد زیادی تحقق نیافته است. در پایان، این مرور نشان می‌دهد که تعامل‌های میان‌رشته‌ای در آینده تا چه اندازه اهمیت خواهند داشت. از یک سو، پژوهشگران علوم انسانی، از جمله علوم انسانی رقمی باید با آرشیوداران

همکاری کنند تا با بهره‌گیری از فنون هوش مصنوعی، راه‌حل‌های آرشیوی متناسب با هدف، طراحی و به کار گیرند. ازسوی دیگر، دانشمندان علوم رایانه و آرشیوداران باید به‌طور مشترک دربارهٔ پیامدهای هوش مصنوعی برای نگهداری اسناد و متقابلاً ظرفیت دانش نگهداری اسناد برای کمک به توسعهٔ داده‌ها و هوش مصنوعی و استفادهٔ مناسب از آن‌ها در جامعه و برای جامعه تأمل کنند. ازاین‌منظر، مرور حاضر برخی حوزه‌های پژوهشی مهم و مرتبط -ازجمله هوش مصنوعی و حقوق و اخلاق هوش مصنوعی - را خارج از دامنهٔ بررسی خود قرار داده است؛ حوزه‌هایی که می‌توانند در پژوهش‌های آینده به‌صورت سودمندی بررسی شوند.

۸. نتیجه‌گیری

در شرایطی که آرشیوداران بیش‌ازپیش به متولی مجموعه‌داده‌هایی با حجم روزافزون تبدیل می‌شوند و هنوز رویکردهای تثبیت‌شده‌ای برای مواجهه با این چالش‌های جدید وجود ندارد، در این مقاله گسترش پژوهش‌ها و آزمایش‌های مرتبط با فناوری‌های هوش مصنوعی در آرشیوها را بررسی کردیم. این مرور، نخستین نمای کلی جامع و متمایز از این مجموعه آثار را ارائه می‌دهد. در این مسیر، رویکردهایی را جست‌وجو و بررسی کردیم که پیش‌تر تأثیری چشمگیر بر رویه‌های موجود نگهداری اسناد داشته‌اند. وجه دیگر این مقاله، روش‌شناسی و ساختاری است که برای سازمان‌دهی آثار مورد بررسی به کار گرفته‌ایم؛ بدین‌ترتیب که آن‌ها را در چارچوب مدل پیوستار اسناد صورت‌بندی کرده‌ایم.

با استفاده از مدل پیوستار اسناد، چهار حوزهٔ متمایز و نیز روندهای نوظهور آینده را شناسایی کردیم. نقطهٔ عزیمت بحث ما این پرسش بود که ادغام هوش مصنوعی چگونه می‌تواند نظری عمل آرشیوی را دگرگون می‌کند. در این زمینه، چندین حوزهٔ همپوشان با نگرانی‌های عام‌تر آرشیوها دربارهٔ تحولات رقمی شناسایی کردیم. فراتر از ملاحظات نظری، روش‌ها و فناوری‌های هوش مصنوعی هم‌اکنون به‌طور فعال برای خودکارسازی بخش‌هایی از جریان‌های کاری آرشیوی، به‌ویژه در ابعاد «ثبت» و «سازمان‌دهی» مدل پیوستار اسناد به کار گرفته می‌شوند.

سپس، فنون نوین هوش مصنوعی مرتبط با ابعاد «سازمان‌دهی» و «تکثر» مدل پیوستار اسناد را بررسی کردیم. ظرفیت فناوری‌های هوش مصنوعی در این زمینه توجه گستردهٔ جوامع آرشیوی را برانگیخته است. ازمنظر سازمان‌دهی، هوش مصنوعی به آرشیوها کمک می‌کند تا از محدودیت‌های واحدهای سنتی آرشیوی فراتر روند و هم‌زمان، نقاط قوت آن‌ها، برای مثال اصل منشأ را حفظ کنند. بخشی از فرصت‌های آینده نیز در انتقال این نقاط قوت به جهان هوش مصنوعی نهفته است. برای مثال، ثبت و حفظ منشأ آرشیوی به فنون خاصی نیاز دارد.

فنون هوش مصنوعی همچنین به‌طور گسترده برای دستیابی به بینش‌های جدید از مجموعه‌هایی به کار گرفته می‌شوند که اکنون رقمی شده‌اند. این فنون، با فراهم‌کردن شکل‌های نوین بازیابی و خوانش دور، به تکثر این مجموعه‌ها کمک می‌کنند و به آرشیوها امکان می‌دهند

مستقیماً با جوامع علمی جدیدی، مانند علوم انسانی رقمی ارتباط برقرار کنند. در نهایت، نقش فناوری‌های هوش مصنوعی را در گشودن راه برای شکل‌های جدید آرشیو، مانند مجموعه‌های رسانه‌های اجتماعی یا دفترهای خاطرات، توضیح دادیم؛ شکل‌هایی که تصور کنونی و نهادمحور ما از آرشیو را گسترش می‌دهند. مقیاس این مجموعه‌ها چالش‌های متمایزی برای سازمان‌دهی و دسترسی ایجاد می‌کند که رفع آن‌ها مستلزم توسعه بیشتر فناوری‌های هوش مصنوعی و ایجاد شیوه‌های دسترسی مستقل‌تر از رابط‌های برنامه‌نویسی کاربردی تجاری است. تنها از این طریق می‌توان مأموریت بنیادین آرشیو، یعنی ایجاد اعتماد به اسناد گذشته را تداوم بخشید.

اگرچه پیشرفت‌های چشمگیری حاصل شده است، همچنان به پژوهش‌های بیشتر و فعالیت مستقیم در آرشیوها و با همکاری آن‌ها نیاز داریم. به‌ویژه، هنوز در تبدیل بسیاری از آزمایش‌ها به رویه‌ها و زیرساخت‌های پایدار آرشیوی شکاف‌هایی وجود دارد. از این رو، پیشنهاد می‌کنیم پژوهش‌های بیشتری بر افزایش اعتماد به فنون هوش مصنوعی، از طریق توسعه چارچوب اخلاقی قوی‌تر و درک بهتر تأثیر این فناوری‌ها بر رویه‌های پژوهشی متمرکز شوند. همچنین، برای روزآمدسازی مبانی نظری عمل آرشیوی با توجه به تحولات اخیر هوش مصنوعی، به پژوهش‌های بیشتری نیاز است. در نهایت، یکی از فرصت‌های مهم، بهره‌گیری از دانش نگهداری اسناد برای کمک به توسعه هوش مصنوعی و استفاده مناسب از آن است؛ دانشی که می‌تواند تخصص مربوط به منشأ، ارزشیابی، زمینه‌مندسازی، شفافیت و پاسخ‌گویی را به جهان داده‌ها وارد کند.

جدول ۱. فهرست مجلات، کنفرانس‌ها و کارگاه‌های بررسی‌شده

Venue	Type	Latest issue
AI & Society	Journal	35(1), September 2020
Archival Science	Journal	20(3), 2020
Archivaria	Journal	89, Spring 2020
Archives and Manuscripts	Journal	48(2), 2020
Archives and Records	Journal	41(1), 2020
Artificial Intelligence and Digital Heritage: Challenges and Opportunities (ARTIDIGH)	Conference/Workshop	2020
Big Data & Society	Journal	7(2), December 2020

Collections: A Journal for Museum and Archives Professionals	Journal	14(1), 2018
COMMA	Journal	2019(1), 2020
Computational Archival Science at IEEE Big Data	Conference/Workshop	2018
Computational Linguistics for Cultural Heritage, Social Sciences, (Humanities and Literature (LaTeCH-CLFL	Conference/Workshop	2019
(Conference on Natural Language Processing (KONVENS	Conference	2020
DH Benelux Journal	Journal	Vol. 2, 2020
(Digital Humanities Quarterly (DHQ	Journal	14(3), 2020
Digital Medievalist	Journal	10, 2015—2016
(Digital Scholarship in the Humanities (DSH	Journal	32(2), June 2020
Digital Studies	Journal	10(1), 2020
Document Analysis Systems	Conference/Workshop	2020
E-Heritage Workshop at ICCV	Conference/Workshop	2019
Fashion, Art, Design Workshop Recurring at ICCV/CVPR	Conference/Workshop	2020
Frontiers in Digital Humanities	Journal	2020
HistoInformatics	Conference/Workshop	2019
Humanist Studies & the Digital Age	Journal	6(1), 2019
Information Discovery and Delivery	Journal	48(3), 2020
International Conference on Frontiers of Handwriting Recognition ((ICFHR	Conference	2018
International Journal of Digital Curation	Journal	15(1), 2020

International Journal of Digital Humanities	Journal	2, 2019
International Journal of Document Analysis and Recognition	Journal	23(3), 2020
International Journal of Humanities and Arts Computing	Journal	14(1—2), 2020
International Journal of Information Management	Journal	52, June 2020
Journal of Archival Organization	Journal	17(1—2), 2020
Journal of Cultural Analytics	Journal	July 2020
Journal of Data Mining and Digital Humanities	Journal	August 2020
Journal of Documentation	Journal	76(5), 2020
Journal of Information Science	Journal	46(4), 2020
Journal of Open Humanities Data	Journal	6, 2020
Journal of the Association for Information Science and Technology (JASIST)	Journal	71(10), 2020
Journal on Computing and Cultural Heritage	Journal	13(3), 2020
Proceedings of the FIAT/IFTA Media Management Seminars	Conference	2017
Records Management Journal	Journal	30(2), 2020
The American Archivist	Journal	82(2), 2019
The Journal of Contemporary Archival Studies	Journal	7, 2020
The Moving Image	Journal	19(1), Spring 2019
Visarts	Conference/Workshop	2020
Visual Pattern Extraction and Recognition for Cultural Heritage (Understanding (VIPERC	Conference/Workshop	2020

References

- Andresen, H. (2020). "A discussion frame for explaining records that are based on algorithmic output." *Records Management Journal*, 30(2), 129-141. <https://doi.org/10.1108/RMJ-04-2019-0019>
- Baron, J. R., & Payne, N. (2017). "Dark archives and e-democracy: Strategies for overcoming access barriers to the public record archives of the future." In *2017 Conference for E-Democracy and Open Government (CeDEM)* (pp. 3-11). IEEE. <https://doi.org/10.1109/CeDEM.2017.27>
- Bell, M. (2020). "From tree to network: Reordering an archival catalogue." *Records Management Journal*, 30(3), 379-394. <https://doi.org/10.1108/RMJ-09-2019-0051>
- Binmakhashen, G. M., & Mahmoud, S. A. (2020). "Document layout analysis: A comprehensive survey." *ACM Computing Surveys*, 52(6), 1-36. <https://doi.org/10.1145/3355610>
- Black, M. L. (2016). "The World Wide Web as complex data set: Expanding the digital humanities into the twentieth century and beyond through internet research." *International Journal of Humanities and Arts Computing*, 10(1), 95-109. <https://doi.org/10.3366/ijhac.2016.0162>
- Blanke, T., & Wilson, J. (2017). "Identifying epochs in text archives." In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE. <https://doi.org/10.1109/BigData.2017.8258172>
- Bocyt, R., & Oomen, J. (2020). "Content adaptation, personalisation and fine-grained retrieval: Applying AI to support engagement with and reuse of archival content at scale." In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*. <https://doi.org/10.5220/0009188505060511>
- Bode, K. (2017). "The equivalence of 'close' and 'distant' reading; or, toward a new object for data-rich literary history." *Modern Language Quarterly*, 78(1), 77-106. <https://doi.org/10.1215/00267929-3699787>
- Bryant, M., Reijnhoudt, L., & Simeonov, B. (2018). "In-place synchronisation of hierarchical archival descriptions." In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE. <https://doi.org/10.1109/BigData.2018.8622517>
- Bunn, J. (2020). "Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI)." *Records Management Journal*, 30(2), 143-153. <https://doi.org/10.1108/RMJ-08-2019-0038>
- Büttner, G. (2019). "Auto-classification in an international organization: Report from a feasibility study." *Comma*, 2017(2), 15-26. <https://doi.org/10.3828/comma.2017.2.2>
- Chabin, M.-A. (2020). "The potential for collaboration between AI and archival science in processing data from the French great national debate." *Records Management Journal*, 30(2), 241-252. <https://doi.org/10.1108/RMJ-08-2019-0042>
- Chaney, A., Wallach, H., Connelly, M., & Blei, D. (2016). "Detecting and characterizing events." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Coeckelbergs, M., & Van Hooland, S. (2020). "Concepts in topics: Using word embeddings to leverage the outcomes of topic modeling for the exploration of digitized archival collections." In R. Mugnaini (Ed.), *Data and information in online environments* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 319). Springer International Publishing. https://doi.org/10.1007/978-3-030-50072-6_4

- Colavizza, G., Ehrmann, M., & Bortoluzzi, F. (2019). "Index-driven digitization and indexation of historical archives." *Frontiers in Digital Humanities*, 6, Article 4. <https://doi.org/10.3389/fdigh.2019.00004>
- Connelly, M. J., Hicks, R., Jervis, R., Spirling, A., & Suong, C. H. (2020). "Diplo-matic documents data for international relations: The Freedom of Information Ar-chive Database." *Conflict Management and Peace Science*. Advance online publi-cation. <https://doi.org/10.1177/0738894220930326>
- de Wilde, M., & Hengchen, S. (2017). "Semantic enrichment of a multilingual ar-chive with linked open data." *Digital Humanities Quarterly*, 11(4).
- Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., & James, S. (2020). "Machine learning for cultural heritage: A survey." *Pattern Recognition Letters*, 133, 102-108. <https://doi.org/10.1016/j.patrec.2020.02.017>
- Gilliland, A. J. (2014). *Conceptualizing 21st-century archives*. Society of Ameri-can Archivists.
- Guldi, J. (2018). "Critical search: A procedure for guided reading in large-scale tex-tual corpora." *Journal of Cultural Analytics*, 3(1). <https://doi.org/10.22148/16.030>
- Gupta, A., & Kapoor, N. (2020). "Comprehensiveness of archives: A modern AI-enabled approach to build comprehensive shared cultural heritage." *In Workshop on Datafication and Cultural Heritage-Provocations, Threats, and Design Op-portunities*. <https://arxiv.org/pdf/2008.04541.pdf>
- Hengchen, S., Coeckelbergs, M., Van Hooland, S., Verborgh, R., & Steiner, T. (2016). "Exploring archives with probabilistic models: Topic modelling for the valorisation of digitised archives of the European Commis-sion." *In 2016 IEEE In-ternational Conference on Big Data (Big Data)*. IEEE. <https://doi.org/10.1109/BigData.2016.7840981>
- Hinrichs, U., Alex, B., Clifford, J., Watson, A., Quigley, A., Klein, E., & Coates, C. M. (2015). "Trading conse-quences: A case study of combining text mining and vis-ualization to facilitate document exploration." *Digital Scholarship in the Humanities*, 30(Suppl. 1), i50-i75. <https://doi.org/10.1093/lc/fqv046>
- Hutchinson, T. (2018). "Protecting privacy in the archives: Supervised machine learning and born-digital records." *In 2018 IEEE International Conference on Big Data (Big Data)*. IEEE. <https://doi.org/10.1109/BigData.2018.8621929>
- Hutchinson, T. (2020). "Natural language processing and machine learning as prac-tical toolsets for archival processing." *Records Management Journal*, 30(2), 155-174. <https://doi.org/10.1108/RMJ-09-2019-0055>
- Jo, E. S., & Geburu, T. (2019). "Lessons from archives: Strategies for collecting so-ciocultural data in machine learning." *In Proceedings of the Conference on Fair-ness, Accountability, and Transparency*. ACM. <https://doi.org/10.1145/3351095.3372829>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks." *In Advances in Neural Information Pro-cessing Systems* (pp. 1097-1105). <http://papers.nips.cc/paper/4824-imagenet-classification-w>
- Lansdall-Welfare, T., & Cristianini, N. (2020). "History playground: A tool for dis-covering temporal trends in massive textual corpora." *Digital Scholarship in the Humanities*, 35(2), 328-341. <https://doi.org/10.1093/lc/fqy077>
- Lee, B. C. G. (2019). "Machine learning, template matching, and the International Tracing Service digital archive: Automating the retrieval of death certificate refer-ence cards from 40 million document scans." *Digital Scholarship in the Humanities*, 34(3), 513-535. <https://doi.org/10.1093/lc/fqy063>
- Lee, C. A. (2018). "Computer-assisted appraisal and selection of archival materials." *In 2018 IEEE International Conference on Big Data (Big Data)*. IEEE. <https://doi.org/10.1109/BigData.2018.8622267>

- Lemieux, V. (2018). "The future of archives as networked, decentralized, auto-nomous and global." In C. Brown (Ed.), *Archival futures* (pp. 33-44). Facet Publishing.
- Lynch, C. (2017). "Stewardship in the 'age of algorithms'." *First Monday*, 22(12). <https://doi.org/10.5210/fin.v22i12.8097>
- Makhoulf Shabou, B., Tièche, J., Knafo, J., & Gaudinat, A. (2020). "Algorithmic methods to explore the automation of the appraisal of structured and unstructured digital data." *Records Management Journal*, 30(2), 175-200. <https://doi.org/10.1108/RMJ-09-2019-0049>
- Marciano, R., Lemieux, V., Hedges, M., Esteva, M., Underwood, W., Kurtz, M., & Conrad, M. (2018). "Archival records and training in the age of big data." In J. Percell, L. C. Sarin, P. T. Jaeger, & J. C. Bertot (Eds.), *Advances in librarianship*. Emerald Publishing Limited. <https://doi.org/10.1108/S0065-28302018000044B010>
- Milligan, I. (2016). "Lost in the infinite archive: The promise and pitfalls of web archives." *International Journal of Humanities and Arts Computing*, 10(1), 78-94. <https://doi.org/10.3366/ijhac.2016.0161>
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). "Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence." *Philosophy & Technology*. Advance online publication. <https://doi.org/10.1007/s13347-020-00405-8>
- Moretti, F. (2013). *Distant reading*. Verso.
- Moretti, G., Sprugnoli, R., Menini, S., & Tonelli, S. (2016). "ALCIDE: Extracting and visualising content from large document collections to support humanities studies." *Knowledge-Based Systems*, 111, 100-112. <https://doi.org/10.1016/j.knosys.2016.08.003>
- Moss, M., & Gollins, T. (2017). "Our digital legacy: An archival perspective." *Journal of Contemporary Archival Studies*, 4(2), Article 3. <https://elischolar.library.yale.edu/jcas/vol4/iss2/3>
- Moss, M., Thomas, D., & Gollins, T. (2018). "The reconfiguration of the archive as data to be mined." *Archivaria*, 86, 118-151.
- Muehlberger, G., Seaward, L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S., Gatos, B., Greinöcker, A., Grüning, T., Hackl, G., Haukovaara, V., Heyer, G., Hirvonen, L., Hodel, T., Jokinen, M., ... Kahle, P. (2019). "Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study." *Journal of Documentation*, 75(5), 954-976. <https://doi.org/10.1108/JD-07-2018-0114>
- Ranade, S. (2016). "Traces through time: A probabilistic approach to connected archival data." In *2016 IEEE International Conference on Big Data (Big Data)*. IEEE. <https://doi.org/10.1109/BigData.2016.7840983>
- Ranade, S. (2018). "Access technologies for the disruptive digital archive." In C. Brown (Ed.), *Archival futures* (pp. 79-98). Facet Publishing.
- Risi, J., Sharma, A., Shah, R., Connelly, M., & Watts, D. J. (2019). "Predicting his-tory." *Nature Human Behaviour*, 3(9), 906-912. <https://doi.org/10.1038/s41562-019-0620-8>
- Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsoupova, T., & Stuart, K. (2019). "More human than human? Artificial intelligence in the archive." *Archives and Manuscripts*, 47(2), 179-203. <https://doi.org/10.1080/01576895.2018.1502088>
- Romein, C. A., Kemman, M., Birkholz, J. M., Baker, J., De Gruijter, M., Meroño-Peñuela, A., Ries, T., Ros, R., & Scagliola, S. (2020). "State of the field: Digital his-tory." *History*, 105(365), 291-312. <https://doi.org/10.1111/1468-229X.12969>
- Ruest, N., Lin, J., Milligan, I., & Fritz, S. (2020). "The Archives Unleashed Project: Technology, process, and

- community to improve scholarly access to web archives." In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. ACM. <https://doi.org/10.1145/3383583.3398513>
- Schneider, J., Adams, C., DeBauche, S., Echols, R., McKean, C., Moran, J., & Waugh, D. (2019). "Appraising, processing, and providing access to email in con-temporary literary archives." *Archives and Manuscripts*, 47(3), 305-326. <https://doi.org/10.1080/01576895.2019.1622138>
- Sødring, T., Reinholdtsen, P., & Massey, D. (2020). "A record-keeping approach to managing IoT-data for government agencies." *Records Management Journal*, 30(2), 221-239. <https://doi.org/10.1108/RMJ-09-2019-0050>
- Sönmez, Ç., Özgür, A., & Yörük, E. (2016). "Towards building a political protest database to explain changes in the welfare state." In *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sci-ences, and Humanities*. <https://doi.org/10.18653/v1/W16-2113>
- Souza, R. R., Coelho, F. C., Shah, R., & Connelly, M. (2016). "Using artificial intel-ligence to identify state secrets." *arXiv*. <https://arxiv.org/abs/1611.00356>
- Spencer, R. (2017). "Binary trees? Automatically identifying the links between born-digital records." *Archives and Manuscripts*, 45(2), 77-99. <https://doi.org/10.1080/01576895.2017.1330158>
- Theimer, K. (2018). "It's the end of the archival profession as we know it, and I feel fine." In C. Brown (Ed.), *Archival futures* (pp. 1-18). Facet Publishing.
- Thibodeau, K. (2016). "Breaking down the invisible wall to enrich archival science and practice." In *2016 IEEE International Conference on Big Data (Big Data)*. IEEE. <https://doi.org/10.1109/BigData.2016.7840986>
- Turchin, P., Brennan, R., Currie, T., Feeney, K., François, P., Hoyer, D., Manning, J., Marciniak, A., Mullins, D., Palmisano, A., Peregrine, P., Turner, E. A. L., & Whitehouse, H. (2015). "Seshat: The global history databank." *Cliodynamics*, 6, 77-107.
- Upward, F., Reed, B., Oliver, G., & Evans, J. (2018). *Recordkeeping informatics for a networked age*. Monash University Publishing.
- Van den Heuvel, C. (2015). "Mapping knowledge exchange in early modern Eu-rope: Intellectual and techno-logical geographies and network representations." *In-ternational Journal of Humanities and Arts Computing*, 9(1), 95-114. <https://doi.org/10.3366/ijhac.2015.0140>
- Vellino, A., & Alberts, I. (2016). "Assisting the appraisal of e-mail records with au-tomatic classification." *Records Management Journal*, 26(3), 293-313. <https://doi.org/10.1108/RMJ-02-2016-0006>
- Weber, M., & Buchanan, R. (2019). "Metadata as a machine for feeling in Germaine Greer's archive." *Archives and Manuscripts*, 47(2), 230-241. <https://doi.org/10.1080/01576895.2019.1568266>
- Wevers, M., & Smits, T. (2019). "The visual digital turn: Using neural networks to study historical images." *Digital Scholarship in the Humanities*, 35(1), 194-207. <https://doi.org/10.1093/llc/fqy085>
- Winters, J., & Prescott, A. (2019). "Negotiating the born-digital: A problem of search." *Archives and Manuscripts*, 47(3), 391-403. <https://doi.org/10.1080/01576895.2019.1568266>