

Studying Comics at Scale with Artificial Intelligence: The PAGE Framework and Three Studies of American Comics and Japanese Manga

Emilio Ferrara^{1*}

^{1*}Thomas Lord Department of Computer Science, University of Southern California, Los Angeles, CA, USA.

Corresponding author(s). E-mail(s): emiliofe@usc.edu;

Abstract

Comics are one of the great visual records of the twentieth century: hundreds of thousands of pages in which a mass readership saw its leaders, its wars and its heroes. Most of this record has never been studied as a whole; comics scholarship proceeds one book at a time. We argue that artificial intelligence, including generative models, can change this to the benefit of the medium itself. We propose PAGE, a framework that makes artificial intelligence an instrument for comics research. It parses every scan into panels, text and a bibliographic record; annotates each page with a vision-language model that follows a written codebook, as a human coder would; grounds samples of these readings in human annotation; and explains change across an archive, with estimates controlled for publisher and genre. We demonstrate the framework on 3,929 American comic books (1938–1956) and 87 volumes of Japanese manga, in three studies, each resting on a different capability. The first pairs text search with visual reading to recover how a mass medium portrayed political power. The second measures the page’s visual form across an industry and recovers the hand of individual artists. The third follows who speaks across two comics traditions, checked against human annotation. Beyond research, the same instruments can make archives searchable by what is drawn, open comics to blind and low-vision readers, and help preserve fragile collections and credit uncredited artists. We outline a research agenda in which artificial intelligence strengthens the study, preservation and reach of the medium.

Keywords: comics, manga, vision-language models, cultural analytics, computational social science, generative AI

1 Introduction

In June 1941, six months before Pearl Harbor, a reader of *Pep Comics* #16 watched a president hand a young hero his commission in the United States Navy. The page never prints the president’s name; the reader was expected to recognize Franklin Roosevelt by his pince-nez and his profile. Hundreds of thousands of such pages survive. A single public collection of golden-age American comic books, the COMICS corpus (Iyyer et al. 2017), holds 3,929 books and 195,795 pages, and the Grand Comics Database (Grand Comics Database 2026; Monaco 2017) indexes 215,961 comic stories in American issues dated 1938 to 1956. Together these pages record how a mass readership saw its leaders, its wars and its heroes, and how a new visual language was invented on cheap newsprint. The questions this record raises concern all of it: how a popular medium reflected and shaped the society that read it, how its visual language formed and spread across publishers, artists and countries, and whose stories it told and whose voices it carried. No scholar can read enough pages to answer them.

Comics scholarship has answered such questions by careful reading of selected books and, more recently, by hand-coded corpora of a few dozen titles (Cohn 2013; Dunst et al. 2018; Pederson and Cohn 2016; Cohn et al. 2017). Computer vision has meanwhile learned to find panels, balloons and characters on a page (Augereau et al. 2018; Nguyen et al. 2018; Sachdeva and Zisserman 2024), and vision-language models can read a page image and answer questions about it in plain language (Vivoli et al. 2024; Baek et al. 2026). These two lines of work have rarely met, because detection models are evaluated on benchmarks and comics scholars seldom have instruments they can trust at the scale of a whole archive. We ask whether artificial intelligence can serve as a trustworthy instrument for studying the comics medium, and whether the indexes, transcripts and attribution evidence it produces can also help the medium’s readers, archives and artists.

We answer with PAGE, a framework that organizes the work into four steps, and with three studies that apply it. *Parse* turns each scan into structured objects: panels, text boxes, reading order, and a bibliographic record with date, publisher and credited artists. *Annotate* has a vision-language model read each page under a written codebook, the same document a human coder would follow, and records what the page shows in the categories the codebook defines, such as how a figure is drawn, what role a character plays and who speaks each line. *Ground* checks samples of those annotations against evidence the model did not produce: a second coder with different information, human annotation where it exists, quantities read by an independent method, and close reading of the pages a finding rests on. *Explain* turns annotations into trends with publisher, genre and sampling held fixed, with intervals that resample comics or volumes, and reads those trends against the history of the medium to answer the study’s question. Built this way, artificial intelligence answers questions about the medium that require reading every page of an archive, and it produces indexes searchable by what is drawn, transcripts that a screen reader can voice and evidence for crediting artists whose names were never printed, all of which can serve the medium’s readers, archives and artists.

Our contributions are four. (i) We define the PAGE framework and catalog the techniques from computer vision, language technology and generative AI that serve

each step, together with uses for readers, archives and artists: discovery, attribution, access, preservation and teaching. (ii) We demonstrate it on 3,929 American comic books and 87 volumes of Japanese manga in three studies, each an instance of one of these questions and each built on a different capability of the framework. To ask how the medium reflected its society, we pair a search of the lettering with a reading of the drawing and recover how golden-age comics portrayed the President of the United States. To ask how its visual language formed, we measure the visual form of every page, trace how the American page changed between 1938 and 1956 and match pages to their credited pencillers. To ask whose voices it carried, we attribute dialogue to speakers and compare who speaks in American comics and in manga. (iii) We show how to ground model annotation in practice, with a second coder working from different information, gender stated by human annotators as a reference for the model’s speaker coding, and checks of dates and panel counts by independent methods. (iv) We set out a research agenda of six lines of work, from discovery by what is drawn to shared human reference sets, in which these instruments serve scholarship and the medium’s readers, archives and artists.

2 Related work

Comics as an object of empirical study.

Comics theory describes a page as a sequence of panels whose meaning arises in the gaps between them (McCloud 1993), and cultural studies read comics as a record of material and everyday culture (Jenkins 2020). Cohn’s account of comics as a visual language, with its own vocabulary and grammar and with distinct American and Japanese dialects, gave the field testable hypotheses (Cohn 2013, 2020, 2021), and empirical comics research tests them on annotated corpora (Dunst et al. 2018; Laubrock and Dunst 2020). Hand-coded studies compare panels in American comics and manga (Cohn 2011; Cohn et al. 2012), trace page layouts and the balance of text and image in American superhero comics across eight decades (Pederson and Cohn 2016; Cohn et al. 2017), and count how much of the dialogue male and female characters speak in manga written for boys and for girls (Unser-Schutz 2015). Annotated collections include eBDtheque (Guérin et al. 2013), the Graphic Narrative Corpus (Dunst et al. 2017), COMICS (Iyyer et al. 2017) and Manga109 (Matsui et al. 2017; Aizawa et al. 2020), with speaker links in Manga109Dialog (Li et al. 2024a) and onomatopoeia in COO (Baek et al. 2022). An early example of large-scale cultural analysis of manga images is Manovich’s visualization of a million fan-translated pages (Manovich et al. 2011), and distant viewing has applied computer vision to comic strips such as *Peanuts* (Wigard et al. 2023).

Computer vision and language technology for comics.

A survey covers panel and balloon detection, text recognition, character recognition and reader interaction (Augereau et al. 2018), and indexing pipelines combine these detectors to make comics searchable (Nguyen et al. 2018). Recent models detect panels, text and characters jointly and associate each balloon with its speaker (Sachdeva and Zisserman 2024; Sachdeva et al. 2024; Li et al. 2024b), translate manga using the

image as context (Hinami et al. 2021), and are evaluated on benchmarks for detection, reading order, speaker identification and narrative closure (Vivoli et al. 2024, 2025; Ikuta et al. 2025; Baek et al. 2026). On the task of choosing the panel that completes a strip, general multimodal models perform near chance (Vivoli et al. 2025), which is one reason we check the model’s annotations against independent evidence before we count them.

Models as measurement instruments for culture.

Quantitative cultural history was popularized by counts of words in digitized books (Michel et al. 2011) and grew into cultural analytics and distant viewing of images (Manovich 2020; Arnold and Tilton 2019; Lee et al. 2020). It now spans narrative and visual media: the emotional arcs of more than a thousand works of fiction reduce to six basic shapes (Reagan et al. 2016), explainable image embeddings trace the aesthetic complexity of 125,000 artworks and recognize their authorship and style (Karjus et al. 2023), and machine learning follows artistic influence across more than 11,000 Japanese Ukiyo-e prints (Honma and Matsui 2026). Computational folkloristics linked large archives of traditional narrative to scholarly categories in searchable graphs (Abello et al. 2012). Computational studies of fiction and film measure who appears and who speaks, finding women under-represented among characters (Underwood et al. 2018; Kraicer and Piper 2019) and portrayed differently in their dialogue (Ramakrishna et al. 2017) and in the power and agency of their actions (Sap et al. 2017); an AI agent, with human-in-the-loop validation, has annotated how female main characters are portrayed in 13,869 novels (Kratzke et al. 2026). Vision-language models have also labeled images at scale, with the labels checked by hand (Chen et al. 2025). Large language models now code text at or above the accuracy of crowd workers (Gilardi et al. 2023; Törnberg 2023), can apply a qualitative codebook at scale (Dunivin 2025), perform best as an aid to human annotators (Ziems et al. 2024), carry demographic and cultural biases (Ferrara 2023), and can change their answers with the surface form of a prompt (Ye et al. 2026). We therefore calibrate a model that reads comics before we use its codes, in the Ground step of our framework.

Generative AI and the people who make comics.

Generative image models raise questions of authorship, labor and consent for artists (Epstein et al. 2023); they have been trained on artists’ work without permission and used to imitate living artists’ styles (Jiang et al. 2023). Zhao and colleagues have built two protections that artists can apply to their own work. Glaze adds small perturbations that cloak an artwork’s style from mimicry by text-to-image models (Shan et al. 2023), and Nightshade turns images scraped for training without consent into poison for the models trained on them, as a defense for content owners (Shan et al. 2024). The same group finds that expert artists and automated detectors each tell human art from AI-generated images well but err in different ways, and argues for combining them (Ha et al. 2024). Synthetic media also threaten the provenance of digital archives (Ferrara 2024, 2026). Generative tools for comics include reference-based colorization (Furusawa et al. 2017) and models that generate manga panels with consistent characters (Wu et al. 2025). We take these concerns as design constraints:

every use of generative AI that we propose serves scholarship, access or preservation, and nothing in this work should be read as an endorsement or enablement of artwork generation with AI.

3 The PAGE framework

PAGE uses a vision-language model as empirical research uses any measurement instrument (Figure 1). A written codebook fixes what the model records before any page is read, and we check a sample of the model’s codes against human annotation and other independent evidence before using them to answer a question about the medium.

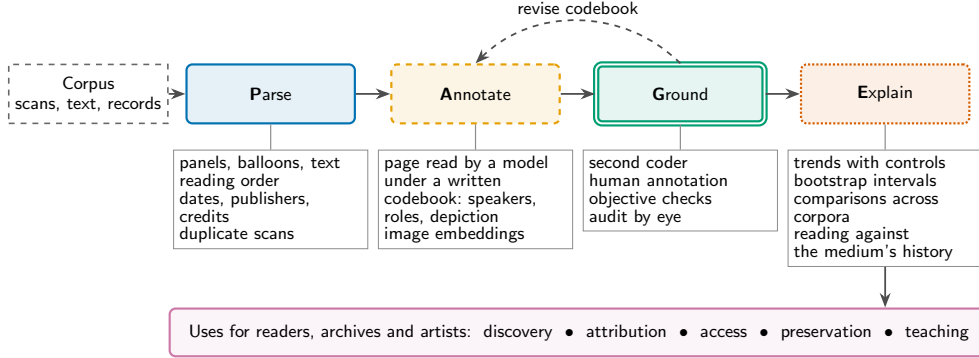


Fig. 1: The PAGE framework. Parse turns scans, text and bibliographic records into structured objects; Annotate has a vision-language model read each page under a written codebook and adds image embeddings; Ground checks samples of the annotations against independent evidence, and disagreements found there lead to revisions of the codebook; Explain estimates trends with publisher, genre and sampling held fixed and reads them against the history of the medium to answer the study’s question. The same objects can serve readers, archives and artists through discovery, attribution, access, preservation and teaching.

Parse.

In the Parse step we detect the panels and text boxes of each page, transcribe the lettering, order the boxes for reading, and attach the page to a bibliographic record. In COMICS, panel and text boxes and an OCR transcription come with the release (Iyyer et al. 2017); for manga we train a Faster R-CNN panel detector (Ren et al. 2017) on the hand-drawn frames of Manga109-s, which reaches box F1 of 0.97 on held-out books, against 0.66 for simple gutter rules. Without a bibliographic record, a page cannot be dated or assigned to a publisher, and public comics corpora often lack these records. The COMICS release carries no dates, publishers or credits, so we read each book’s cover with an on-device text recognizer and match the series and issue printed

in the indicia to the Grand Comics Database (GCD). We also detect the same printed issue scanned twice, so that we count it once.

Annotate.

In the Annotate step we record what each page shows, in categories fixed before the reading. A vision-language model reads the page image, together with any text the parser extracted, under a written codebook that defines each category, its admissible values and its rules for hard cases. We use Claude Opus 5.5 (Anthropic) for every codebook annotation in this paper, run as independent sessions that each read a fixed set of pages. For the manga characters, we discarded a first pass in which some sessions were given more images than they could display, and recoded those characters with one task per session. In the speaker study we shuffle pages across tasks and withhold their years, so that any drift in the model’s coding during a run cannot align with publication year; the model read the presidents pages in batches in corpus order. We also compute CLIP ViT-B/32 image embeddings (Radford et al. 2021), which place each embedded page in a space where nearness means visual similarity and support retrieval, artist attribution and genre classification, as image representations recognize authorship and style in fine art (Karjus et al. 2023).

Ground.

In the Ground step we test whether the model’s annotations are correct against four kinds of evidence that the model did not produce. A second coder recodes a sample from different information, for example the drawings without the character’s name, and we report agreement with Cohen’s κ (Cohen 1960). Where human annotation exists, we score the coding against it directly. Manga109 annotators named many minor characters by their gender (We test the parser with quantities read by an independent method, comparing dates read from the cover by text recognition with dates read from the indicia by the vision model, and detected panels with hand-drawn frames. We check any case a finding rests on, such as a president identified by his face alone, by close reading of the page. We also revise the codebook in response to grounding evidence; the first version of our speaker codebook recorded gender, age and species in one field, and a first coding of 32 pages led us to separate them before the main run.

Explain.

In the Explain step we estimate change in the annotations and read it against what is known of the medium’s history. We make causal claims only where the design supports them. First, every corpus is a selection, because the publishers and genres that survive vary by year, and a trend in the raw means can therefore be a change in the collection. We fit each measure on year with publisher-group and genre fixed effects and report adjusted yearly means. Second, pages of one comic are not independent, so intervals come from a bootstrap that resamples comics or volumes (Efron and Tibshirani 1993). We report contrasts between named windows (for example 1940–41 against 1944–45), so that each claim names the years it compares. We then read the estimates against

the history of the medium, such as the war, the campaign against comics that led to the Comics Code, and the readerships for which manga are published.

Uses for readers, archives and artists.

Table 1 lists the techniques that serve each PAGE step and each use for readers, archives and artists, and marks which ones this paper tests. *Discovery*: archives can become searchable by what is drawn, as Newspaper Navigator made searchable the photographs and illustrations it extracted from sixteen million newspaper pages (Lee et al. 2020). *Attribution*: of the 215,961 American comic stories of 1938–1956 in the GCD, 22% name no penciller, 10% of the stories with a structured penciller credit are flagged as credited in the comic itself, and 8% of credited stories carry a credit marked uncertain; most penciller credits for the period therefore rest on later research. Image embeddings that match pages to their penciller (Section 7) can add a visual signal to that research. *Access*: reading order, speaker links and transcripts, the products of Parse and Annotate, are what a screen reader or an audio description needs for a reader who cannot see the page (Sachdeva and Zisserman 2024), and image-aware translation opens a book to readers of another language (Hinami et al. 2021). *Preservation*: checksummed, indexed scans keep fragile newsprint retrievable, and restored or colorized access copies (Furusawa et al. 2017) can be kept apart from the original scans with their provenance recorded. *Teaching*: pages located by what they show become primary sources for the history of a period, as the presidents study illustrates.

Principles of practice.

Six rules govern the studies below. (1) Every number traces to annotations and every annotation to a page key, so any claim can be checked on the scan. (2) The codebook is written before the main reading and revised only in response to grounding evidence. (3) Grounding uses, where possible, information the model did not have or a reference the model did not produce; Table 2 reports which checks meet this standard. (4) A printed issue scanned twice is counted once wherever it has been detected. (5) We read Manga109-s pages under the dataset’s research license. (6) We document model use as a method, naming the model and keeping its codebooks with the data.

4 Data

COMICS.

The COMICS release provides scans, panel and text boxes, and an OCR transcription with each text box labeled as dialogue or narration, for golden-age American comic books from the Digital Comic Museum (Iyyer et al. 2017). Of its 3,942 books, 3,929 (195,795 pages) carry OCR and panel boxes. We mirror all scans losslessly, recompressing each JPEG to JPEG XL so that every page decodes to a byte-identical copy, and index every page by checksum. Both the scans and the Grand Comics Database index used below are the volunteer work of fans and collectors, part of the participatory culture that has long organized comics fandom (Jenkins 1992, 2006).

Table 1: AI techniques for studying the comics medium, by PAGE step and by use for readers, archives and artists. The last column gives the use in this paper; a dash marks techniques we catalog but did not use.

Step	Task	Technique and example tools	Use in this paper
Parse	Panels and text boxes	Object detection (Faster R-CNN (Ren et al. 2017)); joint panel, text and character detection (Sachdeva and Zisserman 2024)	Panel detector trained on Manga109-s; COMICS release boxes
	Lettering	OCR; on-device text recognition; transcription models (Sachdeva and Zisserman 2024)	COMICS OCR; cover dates
	Bibliographic records	Text recognition on covers and indicia; record linkage to the GCD (Grand Comics Database 2026)	Dates for 77% of comics; GCD publisher, genre, penciller
	Duplicate scans	Perceptual hashing with close inspection	Duplicates among the Study 1 comics
Annotate	Page content under a codebook	Vision-language model reading the page image and extracted text	Presidents, speakers, roles, character gender
	Visual similarity	Image embeddings (CLIP (Radford et al. 2021))	Retrieval, artist matching, cover genre
	Speaker attribution	Speaker detection and character re-identification (Li et al. 2024a,b; Sachdeva et al. 2024)	Human speaker links (Manga109Dialog)
	Text retrieval	Lexical (BM25) and dense sentence embeddings, query expansion	Retrieval comparison (Section 5)
Ground	Agreement	Second coder from different information; Cohen’s κ (Cohen 1960)	Speaker and character coding
	Accuracy	Human annotation as reference; independent methods for dates and counts	Gender-named manga characters; dates; panel frames
Explain	Trends with controls; historical reading	Fixed-effects regression; comic- and volume-level bootstrap (Efron and Tibshirani 1993); interpretation against the history of the medium	All three studies
Use	Discovery	Search by what is drawn (Lee et al. 2020)	–
	Attribution	Nearest-neighbor artist matching on embeddings	Tested on credited pages
	Access	Reading order, speaker links and transcripts for screen readers; image-aware translation (Sachdeva and Zisserman 2024; Hinami et al. 2021)	–
	Preservation	Colorization of access copies (Furusawa et al. 2017); checksummed archives	Lossless, checksummed mirror of all scans
	Teaching	Retrieval of primary-source pages by subject	Presidents study
	Protection	Style cloaking and data poisoning against unauthorized training (Shan et al. 2023, 2024); detection of AI-generated art (Ha et al. 2024)	–

Reconstructed metadata.

The release carries no bibliographic metadata. We read each comic’s cover and first pages with an on-device text recognizer and take the year printed beside a month name or a copyright notice. Compared with the year that the vision model read from the printed cover and indicia of the comics in Study 1, this rule agrees exactly for 139 of 146 comics and within one year for 145. Combining this rule with the scan releases’ own text files and the vision model’s reading of the indicia, we assign a year to 3,034 comics (77% of the release). We match each comic to the GCD (Library of Congress snapshot of 15 April 2026) by the series and issue printed in the indicia, and by year, and link 2,215 comics; on comics whose title and issue we know independently, the matched issue is exact in 72% of cases. The GCD supplies story genres and, story by story, the credited penciller.

Manga.

Manga109-s provides 87 Japanese volumes published from the 1970s to the 2010s, with hand-drawn boxes for every panel, text balloon, face and body (Matsui et al. 2017; Aizawa et al. 2020); the dataset’s catalog gives each title’s decade, readership (shōnen for boys, seinen for men, shōjo for girls, josei for women) and genre. Manga109Dialog links each balloon to the character who speaks it (Li et al. 2024a).

Appendix A documents two properties of the COMICS release that affect any study joining its parts, a page-numbering offset in its panel-annotation subset and duplicate scans of the same printed issue.

5 Grounding the instrument before use

Before we use the model’s codes in the three studies, we check samples of them against human annotation and evidence read by independent methods; Table 2 lists every check by study.

First, the model’s gender coding matches the gender stated by human annotators. Manga109 annotators gave 215 characters names that state their gender (91 of them women), most of them minor characters such as “woman 1” or “mother”. Because the character coding used in Study 3, which codes each manga character once from its crops and its name, saw the names, we recode their crop sheets with the name withheld for this test, and the model gives the annotators’ gender for 95.5% of the 202 characters it classifies as a woman or a man (13 are coded unclear). We also run the page-level procedure that Study 3 applies to American comics (Section 8) on 112 random manga pages and 112 pages chosen to contain such balloons. The model sees each page and its balloons without character names, and it gives the annotators’ gender for the speaker of 92.7% of the 454 balloons spoken by these characters that it codes as a woman or a man (95% CI 85–98%, resampling the 92 speakers; 477 balloons were eligible). The model’s errors go in both directions (17 women coded as men, 16 men coded as women), and on these balloons it gives women 30.4% of the text against 32.6% under the human reference, a difference whose interval includes zero (−14.6 to +6.8 points). Second, codings made under different information or codebook versions agree. In American comics, a first coding of 32 pages under the first codebook version

and the main coding under the revised version agree on the speaker’s gender for 99.6% of the 269 balloons that both coded as a woman or a man ($\kappa = 0.99$); for manga characters, a coder who sees the drawings without the name agrees with the main coder on 96.4% of the 167 of 200 sampled characters that both coded as a woman or a man ($\kappa = 0.93$; 86% exact agreement across all categories). Third, the page-level procedure and the character coding agree on manga pages: on 112 random pages whose balloons come from the human text annotation, the speaker gender assigned on the page matches the coded gender of the linked character for 94.2% of 1,154 balloons ($\kappa = 0.88$), and the two give women 39.5% and 41.6% of the text.

Table 2: Grounding evidence by study. Human references are marked; the other checks compare two readings made with different information or by different methods.

Study	Check	Result
Data	Cover-date OCR rule vs vision-model reading of indicia	139/146 exact
	GCD issue match on known titles	72% exact
	Manga panel detector vs hand-drawn frames (human)	F1 0.97
Page	Cover genre classifier vs GCD genres (human), cross-validated	70% (majority 27%)
	Penciller matching vs GCD credits (human), other series, reprints excluded	2.4% vs 1.5%
Speakers	Character gender, names withheld, vs gender-named characters (human)	95.5%, $n = 202$
	Page procedure vs gender-named speakers (human)	92.7%, $n = 454$
	Two coders, speaker gender (US)	$\kappa = 0.99$, $n = 269$
	Names withheld vs main coder, character gender (manga)	$\kappa = 0.93$, $n = 167$
	Page procedure vs character coding (manga)	94.2%, $\kappa = 0.88$

Lessons from retrieval.

We also apply grounding to the data, in a comparison of text and image retrieval on the panel-annotation subset of COMICS. Claude Opus 5.5 judged the relevance of pages to twelve scene queries (an airplane is drawn, characters fight, a concealed person is revealed) on 102 evaluation pages from 98 comics; one query had no relevant page. On a subset of 45 pages, OpenAI’s **gpt-6-astra** judged the same queries independently, and the two judges agree with $\kappa = 0.58$. Lexical retrieval over raw OCR scores a macro average precision of 0.34, no better than the same index with texts shuffled between pages (0.34) and close to chance (0.32). The cause is a mismatch of vocabulary, because queries written in an analyst’s words, such as Expanding queries into the idiom of the comics and embedding text densely raises text retrieval to 0.47, above the shuffled lexical control by +0.14 (0.09 to 0.20). CLIP page retrieval reaches 0.54, above raw OCR by +0.20 (0.14 to 0.25), and its advantage over dense retrieval with expanded queries is not distinguishable from zero (+0.06, −0.02 to 0.15). An earlier run of the same comparison had joined images to text by the subset’s page keys, which point to the wrong page for 261 of 501 images (Appendix A); the misalignment had overstated the advantage of images.

6 Study 1: how a mass medium portrayed political power

Goal and design.

In this study we pair a search of the lettering with a reading of the drawing, the first capability that PAGE brings to comics research, so that we can ask of an entire archive how a figure is depicted. We ask how American comic books of 1938 to 1956 pictured the President of the United States, and whether the picture changed with the Second World War. We search the OCR of every page for presidential names and titles, which yields 1,005 candidate pages in 684 comics. The vision model reads each candidate page image under a codebook and records who appears, whether he is drawn or only named, how his face is shown, his narrative role and the tone of the treatment. It rejects presidents of companies, clubs and invented republics. Together with 89 sampled pages from lower-precision search terms, and after removing duplicate scans, the model reads 1,082 pages, of which 658 carry a real presidential reference, for 787 references (639 of them dated, a few before 1938). Dates come from the reading checked in Section 5; the portrayal codes have no second coder yet, a limitation we return to in Section 9.

Findings.

First, *sitting presidents are drawn with the face withheld*. Among references that draw the sitting president, 57% withhold his face (95% CI 48–66%, $n = 116$): he appears as a silhouette, in shadow, from the back of the head, or as a voice from behind a desk. Historical presidents are drawn this way in 3% of references (1–7%, $n = 143$); Washington, Lincoln and Theodore Roosevelt appear as full portraits (Figure 2), and part of this group are stamps, statues and wall portraits, whose faces are shown by convention. Several pages show Roosevelt’s face and never print his name (Figure B1). Second, *the commander-in-chief role peaks in wartime*. The share of references that cast the president as commander-in-chief is 4% before Pearl Harbor (2–8%, $n = 182$), 11% in comics dated 1942–45 (7–17%, $n = 137$) and 3% after the war (1–5%, $n = 304$); the wartime share is clearly above the postwar share, and its interval overlaps slightly with the prewar one. Before the war the president is chiefly a national symbol or the authority who dispatches the hero; after it, most references are passing mentions, history features and advertisements for commemorative stamps. Third, *portrayal is deferential*. Of 787 references, 3 mock a real president: a 1921 quip about ex-President Taft in a humor magazine, which falls outside the period, a 1948 gag in which the president hushes an aide to follow a baseball game, and a 1950 caricature. Every other mocking or negative portrayal attaches to a stand-in, such as an impostor dressed as Washington in a Nazi plot or a president hypnotized by a villain.

What PAGE enables.

Text search finds pages that name the president, and the model’s reading of the page image records how he is portrayed. Text alone does not support the first finding, because a portrait and a silhouette carry the same name in the lettering, and pages where Roosevelt is recognizable by his face alone enter the candidate set only when another presidential term appears on them. For this kind of question we therefore pair

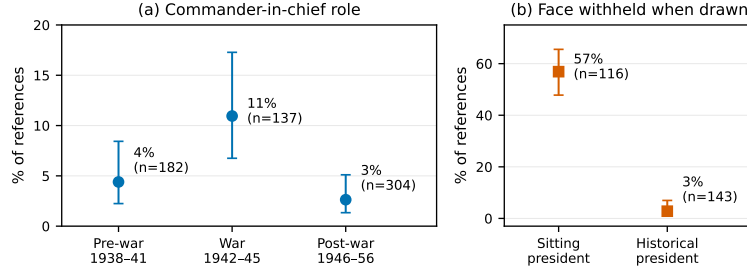


Fig. 2: The President in golden-age comics. Left: share of references in the commander-in-chief role, by publication period. Right: among references that draw a president, the share whose face is withheld, for sitting and historical presidents. Points are proportions with Wilson 95% intervals; n is the number of references, with duplicate scans of one printed issue counted once. Example pages are in Appendix B (Figure B1).

a search of the lettering with the model’s reading of the drawing, and an image search for unnamed likenesses is the next step toward finding pages that name no president.

7 Study 2: the visual form of the page across an industry

Goal and design.

In this study we measure visual form at the scale of an industry. We describe every page by quantities such as panels, splash pages and words, and a fixed sample of pages by an image embedding, so that we can trace changes in the design of the medium, and the work of individual artists, across all surviving books. We ask how the design of the American comic-book page changed between 1938 and 1956, and whether the change holds once we account for which publishers and genres dominate each year. For every page of COMICS we compute, from the release’s panel and text boxes, the number of panels, whether a single panel covers at least 60% of the panel area (a splash), and words of lettering. For each measure we fit, across 2,924 dated comics, a regression on year with publisher-group and genre fixed effects, and report adjusted yearly means with comic-level bootstrap intervals. We compute CLIP image embeddings for a fixed sample of 27,018 pages (the cover and six interior pages per comic). Section 5 checks the cover-genre classifier and the penciller labels against Grand Comics Database records.

Findings.

During the war, pages carried fewer and larger panels. Panels per page fall from 7.28 in 1940–41 to 6.02 in 1944–45, a change of -1.26 panels (95% CI -1.38 to -1.14) with publisher and genre held fixed, and splash pages rise by 4.0 points (3.5 to 4.5). *After the war, words per page rose.* They rise modestly during the war (+21, 14 to 28) and then by 75 words (65 to 85) between 1944–45 and 1953–54, from 189 to 264, and the share of text boxes given to narration returns to its prewar level. Raw and

adjusted curves nearly coincide (Figure 3; typical pages of each period are in Appendix Figure B3), and the contrasts keep their sign and significance when we drop comics dated only through the GCD and when we restrict to comics of known genre, so shifts in the publisher or genre mix of the surviving corpus account for little of either trend. Over eight decades of American superhero comics, hand-coded samples of 40 books show pages moving away from grid layouts and panels carrying fewer words (Pederson and Cohn 2016; Cohn et al. 2017); our whole-corpus estimates cover the first two decades of that span, in which words per page rose after the war.

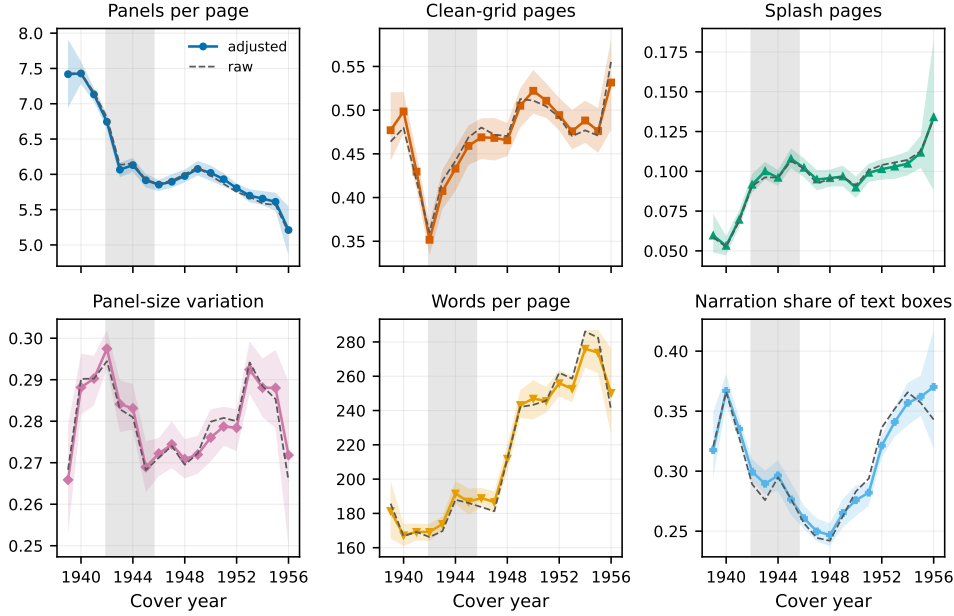


Fig. 3: Page grammar of American comic books by cover year, 1938–1956. Solid lines are adjusted yearly means from comic-level regressions with publisher-group and genre fixed effects, with comic-level bootstrap 95% intervals; dashed lines are raw means. The shaded band marks the United States at war. 1938 is pooled with 1939.

Image embeddings match pages to their penciller. We map sampled pages to stories, and stories to their GCD penciller, which labels 4,497 pages by 351 artists in 1,165 comics. A page’s nearest neighbor among pages of other comics is by the same credited penciller 7.6% of the time, against 2.2% when labels are shuffled within publisher and year. Much of this signal comes from other issues of the same series, which share characters and logos, and from art printed more than once. When we exclude neighbors from the same series, known duplicate scans, reprints of the same art (shared lettering or near-identical images), and every page that nearly duplicates a page of another comic, such as a repeated advertisement, 4,264 pages remain, and the rate is 2.4% against 1.5%, a lift of 1.6 over the shuffled baseline ($p < 0.001$; matched

pairs in Appendix Figure B4). In the network of artists joined to their most similar peers, 32% of links join artists of the same publisher group, against 11% expected by chance, so pages also cluster by publisher group (Figure 4). *Covers record the turn from superheroes to horror*. A linear classifier on cover embeddings, trained on GCD genres, names the genre of a held-out cover correctly 70% of the time (majority class 27%, zero-shot text prompts 46%); applied to every dated cover, superhero covers fall from 42% in 1942–45 to 6% in 1950–56 while horror covers rise from 2% to 26%, the turn that culminated in the campaign against comics (Wertham 1954) and the Comics Code of 1954 (Nyberg 1998; Wright 2001). *Later manga carry fewer panels per page*. In Manga109-s, panels per page fall from 5.3 in 1980s volumes to 5.0 in 2000s volumes and splash pages rise from 11% to 15%, the direction the American page took in the 1940s.

What PAGE enables.

Hand-coded studies of page layout over time rest on a few dozen books (Pederson and Cohn 2016; Cohn et al. 2017), and we extend them to every surviving page, with the composition of the corpus held fixed. The image embeddings also rank pages by their artist, so a corpus of mostly uncredited work can support research on attribution.

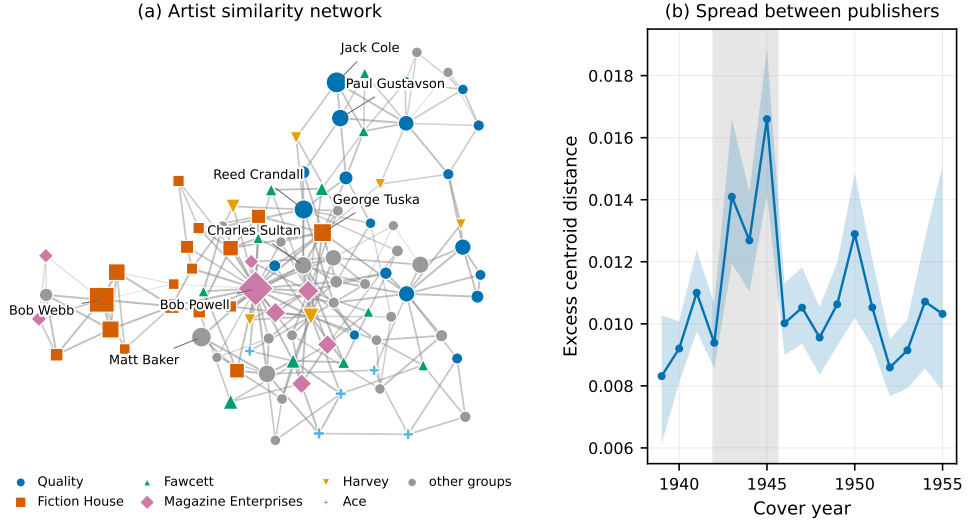


Fig. 4: (a) Artist similarity network: pencillers with at least 15 labeled pages, each joined to its three nearest artists by mean CLIP embedding (Kamada–Kawai layout). Node size gives the artist’s labeled pages; marker shape and color give the main publisher group; the eight artists with the most labeled pages are named. (b) Excess distance between publisher-group centroids by year (observed minus label-shuffled), with 80% subsampling intervals.

8 Study 3: who speaks in two comics traditions

Goal and design.

In this study we attribute dialogue to speakers in two traditions and check the model’s codes against human annotation, so that we can ask the same question about representation of American comics and of Japanese manga. We ask which characters speak the dialogue, and whether the answer differs between the two traditions. We treat each speech balloon as a unit of dialogue with a speaker, an addressee and a word count. For American comics we draw, from each of up to 65 dated comics per year between 1938 and 1956, one interior story page with 3 to 25 dialogue balloons, for 1,132 pages. The vision model reads each page with the list of its dialogue balloons from the OCR and records, for every balloon, the speaker as the character is drawn and presented in the story (gender, age, species), whom the speaker addresses and the speaker’s narrative role, together with the number of women and men drawn on the page. The analysis covers 1,055 story pages and 10,282 speech balloons. For manga, Manga109Dialog identifies the speaker of each balloon, so the model codes each of the 2,684 annotated characters once, from a sheet of crops of that character and the annotators’ name for it. The comparison uses the 7,235 manga pages that match the American sampling frame: interior pages with 3 to 25 text boxes and at least one linked balloon. Japanese is written without spaces, so manga shares are shares of text length in characters. Section 5 shows that the speaker-gender coding gives the gender stated by human annotators for 92.7% of balloons. We exclude speakers coded as groups or as unclear from all shares (example pages from both traditions in Appendix Figures B5 and B6).

Findings.

In American comics, women’s share of the dialogue follows their share of the characters drawn. Women speak 15.7% of the dialogue words, and a woman’s balloon is as long as a man’s (14.2 against 13.7 words). Across most of the period, women’s share of the words on a page moves with their share of the characters drawn, so the low share of dialogue spoken by women arises mainly from the low number of women drawn. *Women’s share of the dialogue rose in the postwar years.* It is 11% in 1938–41 and 10% in 1942–45, rises to 20% in 1946–49 and settles at 18% in 1950–56. Holding publisher and genre fixed, the wartime change (1944–45 against 1940–41) is +2.4 points (95% CI –2.3 to 6.6), and the change from 1940–41 to 1947–49 is +13.3 points (8.5 to 18.5); the share falls back by 1953–54 (Figure 5). In the late 1940s women speak more than their presence alone predicts (adjusted gap +5.0 points above 1940–41, 0.6 to 8.0). *Women mostly address men, and their share of the dialogue is highest on pages with a female protagonist.* Of women’s balloons, 65% address a man and 12% a woman; women speak as the hero’s love interest in 12.9% of their balloons (men 0.9%) and as an authority in 0.7% (men 7.0%). On the 142 pages whose protagonist is coded as a woman, women speak 51% of the words; on the other 913 pages, 10%.

Women’s share of the dialogue is more than twice as high in later manga as in golden-age American comics. Women speak 40.3% of the dialogue in manga (95% CI 34.9 to 45.6), against 15.7% (14.2 to 17.3) in American comics of 1938–56 (Figure 6). The difference holds within every readership: 34.7% in shōnen manga written for boys,

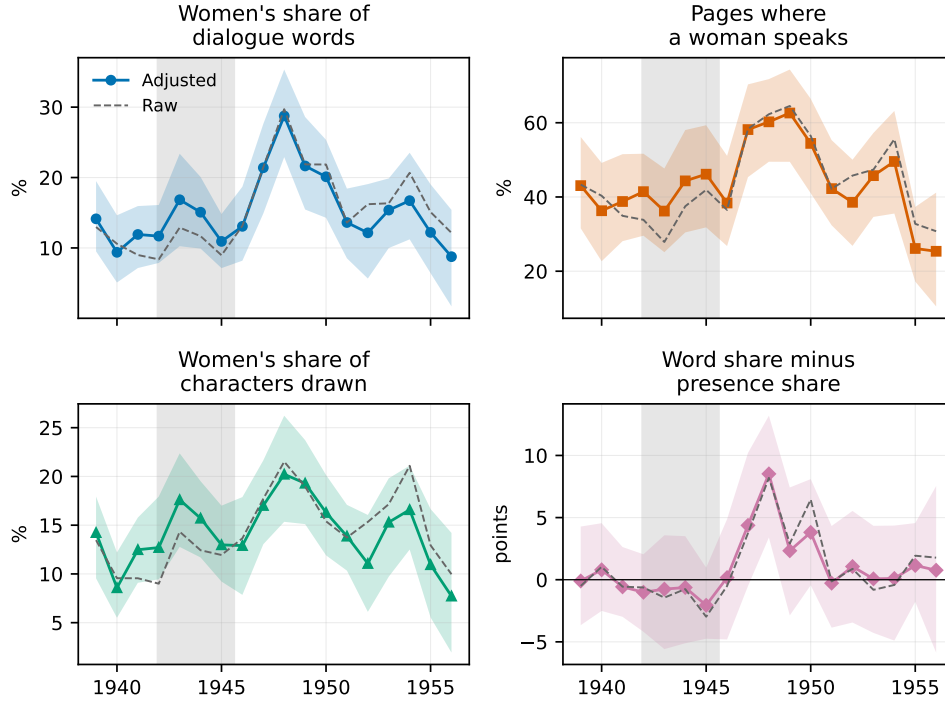


Fig. 5: Women’s voice and presence in American comic books, 1938–1956. Solid lines are yearly estimates adjusted for publisher group and page genre, with comic-bootstrap 95% bands; dashed lines are raw yearly means. The gray band marks the United States at war. Lower right: women’s share of dialogue words minus their share of the characters drawn on the page; zero means women speak in proportion to their presence.

39.2% in seinen for men, 41.7% in josei for women and 49.4% in shōjo for girls. A study of ten series found that male characters speak most of the dialogue in shōnen manga and that shōjo is more balanced (Unser-Schutz 2015), and our estimates extend that finding to four readerships and a whole collection. The two corpora are separated by roughly half a century, American comics of 1938–1956 against manga of the 1970s to the 2010s, so the contrast mixes the difference between the two traditions with the difference between periods; American comics of the same later decades may well be more balanced, which a comparison of matched periods would test. As in American comics, women’s share of the dialogue largely tracks their share of the characters on the page (gap -0.3 points in manga, 1.0 in American comics), so the difference between the two corpora lies mainly in how many women they draw. In shōjo manga, by contrast, women speak more than their presence predicts (4.0 points, 1.2 to 6.6).

What PAGE enables.

Human speaker links exist for one manga collection and for no American archive, so we use the model to attribute speakers in American comics and the human annotation of manga to check the model’s codes, which lets us compare the two traditions with one instrument.

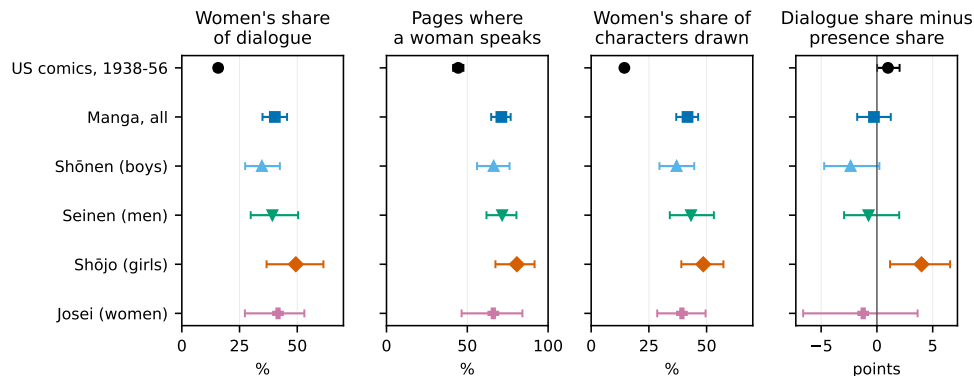


Fig. 6: Women’s voice and presence in American comic books (1938–1956, one random story page with 3 to 25 dialogue balloons per comic, word shares) and Japanese manga (Manga109-s, 1970s–2010s, 7,235 interior pages with 3 to 25 text boxes and at least one human speaker link, shares of text length), overall and by the readership a manga is published for. Points are estimates with 95% bootstrap intervals, resampling comics for the American sample and volumes for manga. Right: women’s share of the dialogue minus their share of the characters drawn. Manga speakers come from the Manga109Dialog human speaker links.

9 A research agenda for artificial intelligence and comics

The withheld face of the sitting president, the wartime change of the page, and the relation between women’s dialogue and women’s presence each rest on a different capability of PAGE and on thousands of pages, and none could be established without reading every page of an archive. The same instruments open six lines of work for scholarship and for the medium’s readers, archives and artists, and for each we state what this page already provides.

(1) *Discovery by what is drawn.*

Archives of comics are indexed by title, issue and, at best, the words on the page. Image embeddings and vision-language reading make them searchable by what is drawn, as Newspaper Navigator did for the photographs and illustrations of historical newspapers (Lee et al. 2020). Study 1 shows the need, because text search reaches pages

on which Roosevelt is recognizable only by his face only when another presidential term happens to appear on them. An image search for unnamed likenesses would let a teacher or a historian assemble every such page.

(2) Attribution of uncredited artists.

Of the 215,961 American comic stories of 1938–1956 in the GCD, 22% name no penciller, and most of the rest carry credits established by later research. Across series, the nearest page is by the same credited penciller for 2.4% of pages, 1.6 times the shuffled baseline. Embeddings therefore add a visual signal that, combined with signatures, series records and the judgment of experts, can help credit the artists of the period. The open technical problem is to separate an artist’s drawing from recurring characters, which can run in more than one series, and from repeated art.

(3) Access for readers who cannot see the page.

A screen reader or an audio description needs the reading order, speaker links and transcripts that Parse and Annotate already produce (Sachdeva and Zisserman 2024). Generative models can describe what a panel shows, and image-aware translation opens a book to readers of another language (Hinami et al. 2021). We propose to build and evaluate these tools with blind and low-vision readers.

(4) Preservation and restoration with provenance.

The checksummed, indexed mirror we built for this study already keeps the COMICS scans retrievable. Generative restoration and colorization (Furusawa et al. 2017) can produce readable access copies of damaged pages, provided each copy is kept apart from the original scan and labeled with how it was made. Archives will also need to detect synthetic pages that enter their collections (Ferrara 2026), a task in which expert readers and automated detectors work best together (Ha et al. 2024).

(5) A comparative science of the medium.

Theories of comics as a visual language with national dialects (Cohn 2013, 2011) become testable across whole traditions once pages from several of them are parsed and annotated with one instrument. Study 3 compares American comics and manga; extending the comparison to matched periods, and to European comics and webcomics, would turn the present contrasts into tests of how traditions diverge and influence one another.

(6) Shared instruments and human references.

Model readings become cumulative evidence when their codebooks, annotations and human references are shared. The gender-named characters of Manga109 served here as a human reference that required no new annotation; deliberate reference sets for narrative role, depiction and speaker identity, and replication of each codebook across model families, would let every future study report its error against the same standard. The fan communities that assembled these archives are natural partners in building such references (Jenkins 2006).

Risks and how the framework measures them.

A model that reads comics can make systematic errors that two of its own sessions will reproduce. It can carry the demographic and cultural biases of its training (Ferrara 2023) and change its answers with the wording of a prompt (Ye et al. 2026). We measure these errors on samples checked against independent evidence (Section 5), including the human reference for speaker gender in manga. Generative AI also poses risks to the people who make comics, through imitation of their styles and training on their work without consent (Jiang et al. 2023; Epstein et al. 2023), and to archives, through synthetic pages that blur provenance (Ferrara 2026); artists can protect their work against the first with style cloaking and data poisoning (Shan et al. 2023, 2024). Every use we propose is analytic or serves access, none generates new pages in an artist’s style, and attribution remains the decision of human experts.

Limitations.

Gender is coded as each character is drawn and presented in the story, not as any real identity. The presidents study has no second coder; the same model family reviewed its likeness-based identifications in a second reading of the page image, these identifications await a human check, and its discovery reaches only pages whose OCR names a president. All codebook annotation uses one model family, and the human references cover speaker gender in manga, mostly for minor characters; narrative role, addressee and tone have no human reference, nor do the American pages, whose balloons come from noisy OCR and whose speaker sample draws one random page per comic. Duplicate scans were searched only among the comics of Study 1, and selection within publishers and genres cannot be ruled out by fixed effects. The manga comparison spans different decades (the 1970s to the 2010s against 1938–1956), and Manga109-s is a curated set of 87 volumes, so we draw no causal conclusion from the contrast between the two traditions. Penciller matching cannot exclude recurring features that run in more than one series. COMICS is a selection of public-domain books, and its reconstructed dates and GCD links carry noise that biases tests such as penciller matching toward the null.

10 Conclusions

Comics are a visual record that a mass readership lived with for generations, and most of that record has never been studied as a whole. We proposed PAGE, a framework that parses pages into structured objects, annotates them with a vision-language model under written codebooks, grounds samples of the model’s codes in human annotation and independent checks, and explains change across an archive with publisher and genre held fixed. Three studies of 3,929 American comic books and 87 volumes of manga, each built on a different capability, find that golden-age comics withheld the sitting president’s face, that the American page carried fewer and larger panels in wartime and more words after it, and that in both traditions women’s share of the dialogue largely tracks their share of the characters drawn. The same instruments can make archives searchable by what is drawn, open comics to blind and low-vision readers, help preserve fragile collections and help credit the artists of the period.

Grounded in human annotation, and used with the consent and credit of the people who make comics, artificial intelligence can extend the study of the medium to whole archives and serve its readers, archives and artists.

Acknowledgements. We thank the creators of the COMICS corpus, the Digital Comic Museum, the Grand Comics Database volunteers, and the Manga109 and Manga109Dialog teams at the University of Tokyo. Manga109-s is used under its research license, which permits reproducing up to 20% of the pages of each volume with credit to the author; the four manga pages in Appendix Figure B6 follow these terms.

Declarations

- **Availability of data and materials.** COMICS is available from its authors (Iyyer et al. 2017); Manga109-s and Manga109Dialog are available from the University of Tokyo under their licenses (Aizawa et al. 2020; Li et al. 2024a); the GCD is available at <https://www.comics.org> (Grand Comics Database 2026). The code, codebooks, annotations and derived measures will be released in a public repository upon publication; page images are not redistributed.
- **Use of large language models.** Claude Opus 5.5 (Anthropic) was the annotation instrument described in Section 3: it coded the presidents pages, the speaker pages and the manga characters, read publication years and titles from covers and indicia, and judged retrieval relevance in Section 5 (OpenAI’s gpt-6-astra, run through the Codex CLI, judged a subset of 45 pages, with $\kappa = 0.58$ between the two). Claude also assisted with code, reference retrieval and drafting under the author’s direction; the author reviewed all outputs and takes responsibility for the content.
- **Ethics approval and consent to participate.** Not applicable.
- **Competing interests.** The author declares no competing interests.
- **Funding.** Not applicable.
- **Authors’ contributions.** EF conceived the study, designed the framework, directed the analyses and wrote the manuscript.

Appendix A Data notes

Page numbering in the panel-annotation subset of COMICS.

The widely used panel-annotation subset numbers its page images differently from the full archive. When we match all 501 images by checksum to the byte-identical page in the full scans, 261 carry a different page key, almost always one page higher. The OCR follows the full archive, so a study that joins the subset’s images to the OCR by key reads the wrong page’s text for about half the pages; all results in this article use the full archive.

Duplicate scans.

Some printed issues appear twice in COMICS as separate scans (for example *Liberty Scouts Comics #2* as comics 1744 and 1745). Among the 684 comics read for Study 1 we find 4 such pairs by perceptual hash, confirm each by inspection, and count each

issue once in every study. Pages that reprint the same art in another series, and advertisements repeated across comics, are excluded from the penciller-matching test in Study 2 (Appendix B, Figure B4).

Appendix B Gallery of pages

COMICS pages in this gallery are public-domain scans from the Digital Comic Museum, reproduced unaltered and identified by their page key in the COMICS release. Manga pages are from Manga109-s, reproduced under its license (one page per volume, courtesy of the author) (Matsui et al. 2017; Aizawa et al. 2020), with speakers from Manga109Dialog (Li et al. 2024a).



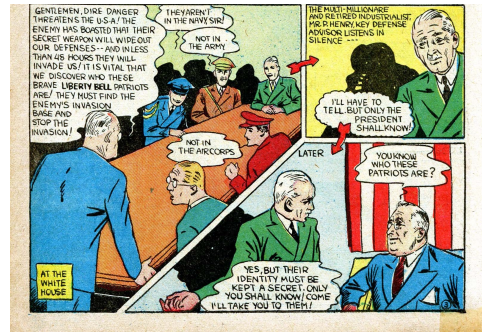
(a) *Pep Comics* #16 (M.L.J., June 1941), 2124-45. The president presenting Navy commissions is never named; the pince-nez and the features point to Franklin Roosevelt.



(b) *Smash Comics* #6 (Quality, January 1940), 2630-62. An unnamed, “weary-voiced” president pledges neutrality at the microphone.



(c) *Big 3* #2 (Fox, Winter 1941), 383-59. The recovered president demands that “this fiend must be found”.



(d) *Liberty Scouts Comics* #2 (Comic Corporation of America, June 1941), 1745-4. The president, seen from behind at the White House and then at his desk, is never named.

Fig. B1: Study 1: the President in golden-age comics. Unaltered crops of the original scans; titles and dates are read from each comic’s cover and indicia.



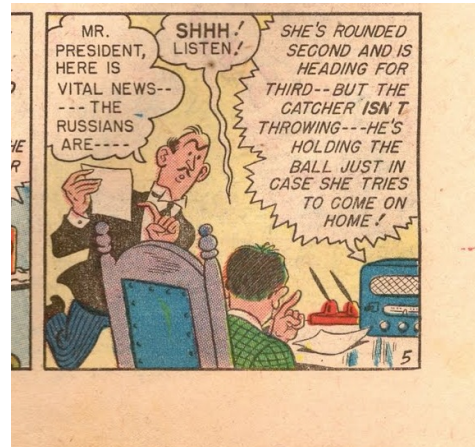
(a) *War Heroes* #4 (Dell, April–June 1943), 3150.4. The president, unnamed, with Generals Clark and Eisenhower; the caption says only “high government executives”.



(b) *Dynamic Comics* #3 (Chesler, year not printed), 1039.34. Churchill and the president after a foiled assassination.



(c) *Police Comics* #15 (Quality, January 1943), 2265.10. A Nazi shouts “Down with Roosevelt!”; the face is most likely Plastic Man’s disguise.



(d) *Babe* #3 (Prize, October–November 1948), 293.36. One of the three references that mock a real president: he hushes an aide to follow a baseball game.

Fig. B2: Study 1: the President in golden-age comics (continued). Unaltered crops of the original scans; titles and dates are read from each comic’s cover and indicia.

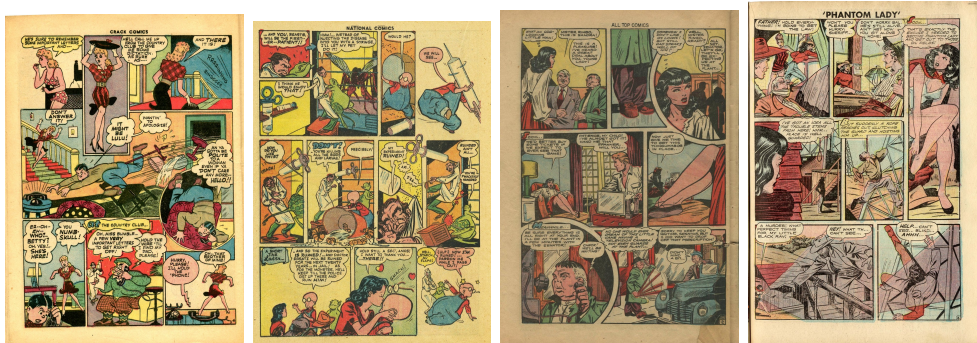


(a) 1940–41: *Whiz Comics* #18, Fawcett, 1941, 3934-35; 8 panels, 142 words. (b) 1940–41: Title not matched, 1941, 2903-15; 8 panels, 148 words. (c) 1944–45: *Captain Aero Comics* #3 [17], 1944, 624-46; 7 panels, 158 words.

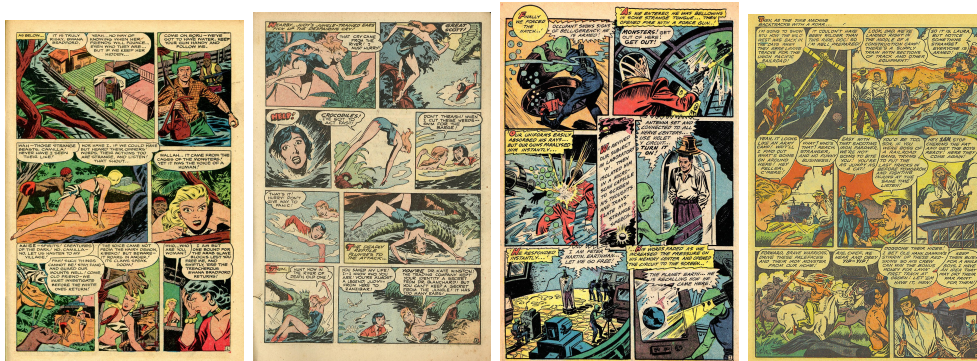


(d) 1944–45: *National Comics* #49, Quality, 1945, 1996-10; 7 panels, 161 words. (e) 1953–54: *Out of the Night* #16, ACG, 1954, 80-7; 7 panels, 195 words. (f) 1953–54: *Trapped!* #1, Ace, 1954, 3085-34; 7 panels, 196 words.

Fig. B3: Study 2: typical pages before, during and after the war. In each window, the two story pages of the speaker sample closest to the window's median panels and words per page (medians 1940–41: 8 panels and 142 words; 1944–45: 7 panels and 157 words; 1953–54: 7 panels and 195 words).

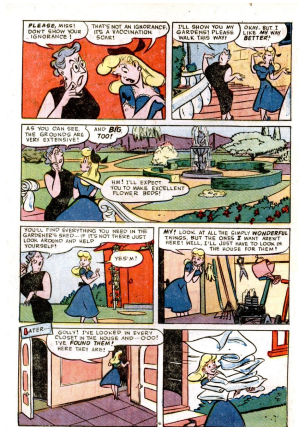


(a) *Crack Comics* #32, Quality, 1943, 3496_35. (b) *National Comics* #50, Quality, 1945, 2424_35. (c) *All Top Comics* #10, Fox, 1948, 163_27. (d) *Phantom Lady* #20, Fox, 1948, 2173_14.



(e) *Jungle Comics* #131, Fiction House, 1950, 3683_40. (f) *Exciting Comics* #68, Fiction House, 1952, Nedor/Standard, 8_9. (g) *Ghost Comics* #4, 1949, 1088_14. (h) *Jumbo Comics* #131, Fiction House, 1949, 1580_21.

Fig. B4: Study 2: pages matched to their nearest neighbor in another series by the same credited penciller: Bernard Dibble (a–b), Matt Baker (c–d), Ralph Mayo (e–f), Jack Kamen (g–h). Pairs are the most similar such matches after excluding reprints and repeated pages (image cosine ≥ 0.95 to a page of another comic, or shared lettering).

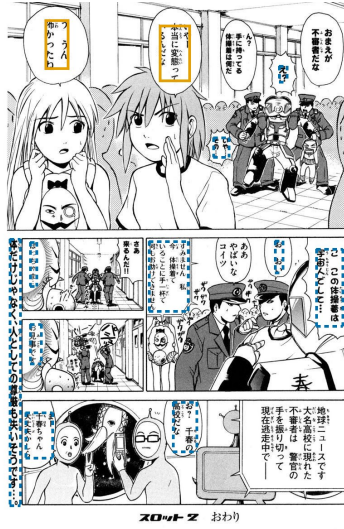


(a) Title not matched, 1955, 990_7: the female-led page with the most balloons spoken by women. Coded speakers: 11 balloons by women, 0 by men; women speak 100% of the words; drawn: 2 women, 0 men.



(b) *Power Comics* #2, 1945, 2085_58: the page closest to the median women's word share among pages that draw a woman. Coded speakers: 4 balloons by women, 13 by men; women speak 23% of the words; drawn: 1 woman, 4 men.

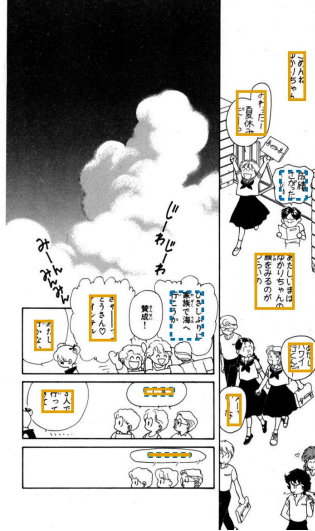
Fig. B5: Study 3: American story pages from the speaker sample, with the model's codes.



(a) shōnen (boys): ランスロットフルスロットル, 2000s; courtesy of ダイナミック 太郎. Linked balloons: 2 by women, 10 by men.



(b) seinen (men): チャンギリぼんぼん, 1980s; courtesy of 木村 直巳. Linked balloons: 3 by women, 5 by men.



(c) shōjo (girls): その気でA B C, 1980s; courtesy of たしろ 季霧. Linked balloons: 10 by women, 4 by men.



(d) josei (women): うちの猫ず日記, 2000s; courtesy of があさん. Linked balloons: 5 by women, 8 by men.

Fig. B6: Study 3: one Manga109-s page per readership, with each balloon outlined by the coded gender of the speaker that Manga109Dialog annotators link to it (orange solid outline: woman; blue dashed outline: man). Each is the page with the most linked balloons, among the manga pages of its readership coded with the page-level procedure in Section 5, on which both women and men speak.

References

- Abello J, Broadwell P, Tangherlini TR (2012) Computational folkloristics. *Communications of the ACM* 55(7):60–70. <https://doi.org/10.1145/2209249.2209267>
- Aizawa K, Fujimoto A, Otsubo A, et al (2020) Building a Manga Dataset “Manga109” With Annotations for Multimedia Applications. *IEEE MultiMedia* 27(2):8–18. <https://doi.org/10.1109/mmul.2020.2987895>
- Arnold T, Tilton L (2019) Distant viewing: analyzing large visual corpora. *Digital Scholarship in the Humanities* 34(Supplement_1):i3–i16. <https://doi.org/10.1093/llc/fqz013>
- Augereau O, Iwata M, Kise K (2018) A Survey of Comics Research in Computer Science. *Journal of Imaging* 4(7):87. <https://doi.org/10.3390/jimaging4070087>
- Baek J, Matsui Y, Aizawa K (2022) COO: Comic Onomatopoeia Dataset for Recognizing Arbitrary or Truncated Texts. In: *Computer Vision – ECCV 2022*. Springer Nature Switzerland, pp 267–283, https://doi.org/10.1007/978-3-031-19815-1_16
- Baek J, Egashira K, Onohara S, et al (2026) MangaVQA and MangaLMM: A Benchmark and Specialized Model for Multimodal Manga Understanding. In: *Findings of the Association for Computational Linguistics: EACL 2026*. Association for Computational Linguistics, pp 5349–5370, <https://doi.org/10.18653/v1/2026.findings-eacl.284>
- Chen Z, Ye J, Tsai B, et al (2025) Synthetic Politics: Prevalence, Spreaders, and Emotional Reception of AI-Generated Political Images on X. In: *Proceedings of the 36th ACM Conference on Hypertext and Social Media*. ACM, pp 11–21, <https://doi.org/10.1145/3720553.3746675>
- Cohen J (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20(1):37–46. <https://doi.org/10.1177/001316446002000104>
- Cohn N (2011) A different kind of cultural frame: An analysis of panels in American comics and Japanese manga. *Image & Narrative* 12(1):120–134. URL <https://www.imageandnarrative.be/index.php/imagenarrative/article/view/128>
- Cohn N (2013) *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. Bloomsbury Academic, <https://doi.org/10.5040/9781472542175>
- Cohn N (2020) Your Brain on Comics: A Cognitive Model of Visual Narrative Comprehension. *Topics in Cognitive Science* 12(1):352–386. <https://doi.org/10.1111/tops.12421>

- Cohn N (2021) Who Understands Comics?: Questioning the Universality of Visual Language Comprehension. Bloomsbury Academic, <https://doi.org/10.5040/9781350156074>
- Cohn N, Taylor-Weiner A, Grossman S (2012) Framing Attention in Japanese and American Comics: Cross-Cultural Differences in Attentional Structure. *Frontiers in Psychology* 3:349. <https://doi.org/10.3389/fpsyg.2012.00349>
- Cohn N, Taylor R, Pederson K (2017) A Picture is Worth More Words Over Time: Multimodality and Narrative Structure Across Eight Decades of American Superhero Comics. *Multimodal Communication* 6(1):19–37. <https://doi.org/10.1515/mc-2017-0003>
- Dunivin ZO (2025) Scaling hermeneutics: a guide to qualitative coding with LLMs for reflexive content analysis. *EPJ Data Science* 14:28. <https://doi.org/10.1140/epjds/s13688-025-00548-8>
- Dunst A, Hartel R, Laubrock J (2017) The Graphic Narrative Corpus (GNC): Design, Annotation, and Analysis for the Digital Humanities. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). IEEE, pp 15–20, <https://doi.org/10.1109/icdar.2017.286>
- Dunst A, Laubrock J, Wildfeuer J (eds) (2018) Empirical Comics Research: Digital, Multimodal, and Cognitive Methods. Routledge, <https://doi.org/10.4324/9781315185354>
- Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Springer US, <https://doi.org/10.1007/978-1-4899-4541-9>
- Epstein Z, Hertzmann A, the Investigators of Human Creativity, et al (2023) Art and the science of generative AI. *Science* 380(6650):1110–1111. <https://doi.org/10.1126/science.adh4451>
- Ferrara E (2023) Should ChatGPT be biased? Challenges and risks of bias in large language models. *First Monday* 28(11). <https://doi.org/10.5210/fm.v28i11.13346>
- Ferrara E (2024) GenAI against humanity: nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science* 7(1):549–569. <https://doi.org/10.1007/s42001-024-00250-1>
- Ferrara E (2026) The Generative AI Paradox: GenAI and the Erosion of Trust, the Corrosion of Information Verification, and the Demise of Truth. *Future Internet* 18(2):73. <https://doi.org/10.3390/fi18020073>
- Furusawa C, Hiroshiba K, Ogaki K, et al (2017) Comicolorization: semi-automatic manga colorization. In: SIGGRAPH Asia 2017 Technical Briefs. ACM, pp 1–4, <https://doi.org/10.1145/3145749.3149430>

- Gilardi F, Alizadeh M, Kubli M (2023) ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences* 120(30):e2305016120. <https://doi.org/10.1073/pnas.2305016120>
- Grand Comics Database (2026) Grand Comics Database. URL <https://www.comics.org>, data snapshot of 15 April 2026
- Guérin C, Rigaud C, Mercier A, et al (2013) eBDtheque: A Representative Database of Comics. In: 2013 12th International Conference on Document Analysis and Recognition. IEEE, pp 1145–1149, <https://doi.org/10.1109/icdar.2013.232>
- Ha AYJ, Passananti J, Bhaskar R, et al (2024) Organic or Diffused: Can We Distinguish Human Art from AI-generated Images? In: *Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security*. ACM, pp 4822–4836, <https://doi.org/10.1145/3658644.3670306>
- Hinami R, Ishiwatari S, Yasuda K, et al (2021) Towards Fully Automated Manga Translation. *Proceedings of the AAAI Conference on Artificial Intelligence* 35(14):12998–13008. <https://doi.org/10.1609/aaai.v35i14.17537>
- Honna S, Matsui A (2026) Disruptive transformation of artworks in master-disciple relationships: the case of Ukiyo-e artworks. *EPJ Data Science* 15:45. <https://doi.org/10.1140/epjds/s13688-026-00646-1>
- Ikuta H, Wöhler L, Aizawa K (2025) MangaUB: A Manga Understanding Benchmark for Large Multimodal Models. *IEEE MultiMedia* 32(2):33–43. <https://doi.org/10.1109/mmul.2025.3550451>
- Iyyer M, Manjunatha V, Guha A, et al (2017) The Amazing Mysteries of the Gutter: Drawing Inferences Between Panels in Comic Book Narratives. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp 6478–6487, <https://doi.org/10.1109/cvpr.2017.686>
- Jenkins H (1992) *Textual Poachers: Television Fans and Participatory Culture*. Routledge, New York
- Jenkins H (2006) *Convergence Culture: Where Old and New Media Collide*. New York University Press, New York
- Jenkins H (2020) *Comics and Stuff*. New York University Press, New York, <https://doi.org/10.18574/nyu/9781479831258.001.0001>
- Jiang HH, Brown L, Cheng J, et al (2023) AI Art and its Impact on Artists. In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, pp 363–374, <https://doi.org/10.1145/3600211.3604681>

- Karjus A, Canet Solà M, Ohm T, et al (2023) Compression ensembles quantify aesthetic complexity and the evolution of visual art. *EPJ Data Science* 12:21. <https://doi.org/10.1140/epjds/s13688-023-00397-3>
- Kraicer E, Piper A (2019) Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction. *Journal of Cultural Analytics* 3(2). <https://doi.org/10.22148/16.032>
- Kratzke S, Meral E, Rosenbusch H (2026) Do female and male authors differ in their portrayal of female characters? An AI-assisted exploration of public book records across genres and time. *EPJ Data Science* 15:66. <https://doi.org/10.1140/epjds/s13688-026-00665-y>
- Laubrock J, Dunst A (2020) Computational Approaches to Comics Analysis. *Topics in Cognitive Science* 12(1):274–310. <https://doi.org/10.1111/tops.12476>
- Lee BCG, Mears J, Jakeway E, et al (2020) The Newspaper Navigator Dataset: Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. ACM, pp 3055–3062, <https://doi.org/10.1145/3340531.3412767>
- Li Y, Aizawa K, Matsui Y (2024a) Manga109Dialog: A Large-Scale Dialogue Dataset for Comics Speaker Detection. In: *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, pp 1–6, <https://doi.org/10.1109/icme57554.2024.10687709>
- Li Y, Hinami R, Aizawa K, et al (2024b) Zero-Shot Character Identification and Speaker Prediction in Comics via Iterative Multimodal Fusion. In: *Proceedings of the 32nd ACM International Conference on Multimedia*. ACM, pp 7366–7374, <https://doi.org/10.1145/3664647.3681659>
- Manovich L (2020) *Cultural Analytics*. The MIT Press, <https://doi.org/10.7551/mitpress/11214.001.0001>
- Manovich L, Douglass J, Huber W (2011) Understanding scanlation: how to read one million fan-translated manga pages. *Image & Narrative* 12(1):206–228. URL <https://doaj.org/article/90d0806623ef4b9392915c85e69171cb>
- Matsui Y, Ito K, Aramaki Y, et al (2017) Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications* 76(20):21811–21838. <https://doi.org/10.1007/s11042-016-4020-z>
- McCloud S (1993) *Understanding Comics: The Invisible Art*. Kitchen Sink Press, Northampton, MA

- Michel JB, Shen YK, Aiden AP, et al (2011) Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331(6014):176–182. <https://doi.org/10.1126/science.1199644>
- Monaco M (2017) Grand Comics Database: <http://www.comics.org/>. *Technical Services Quarterly* 34(1):108–109. <https://doi.org/10.1080/07317131.2017.1238215>
- Nguyen NV, Rigaud C, Burie JC (2018) Digital Comics Image Indexing Based on Deep Learning. *Journal of Imaging* 4(7):89. <https://doi.org/10.3390/jimaging4070089>
- Nyberg AK (1998) *Seal of Approval: The History of the Comics Code*. University Press of Mississippi, Jackson, MS
- Pederson K, Cohn N (2016) The changing pages of comics: Page layouts across eight decades of American superhero comics. *Studies in Comics* 7(1):7–28. https://doi.org/10.1386/stic.7.1.7_1
- Radford A, Kim JW, Hallacy C, et al (2021) Learning Transferable Visual Models From Natural Language Supervision. In: *Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research*, vol 139. PMLR, pp 8748–8763, URL <https://proceedings.mlr.press/v139/radford21a.html>
- Ramakrishna A, Martínez VR, Malandrakis N, et al (2017) Linguistic analysis of differences in portrayal of movie characters. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pp 1669–1678, <https://doi.org/10.18653/v1/p17-1153>
- Reagan AJ, Mitchell L, Kiley D, et al (2016) The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science* 5:31. <https://doi.org/10.1140/epjds/s13688-016-0093-1>
- Ren S, He K, Girshick R, et al (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(6):1137–1149. <https://doi.org/10.1109/tpami.2016.2577031>
- Sachdeva R, Zisserman A (2024) The Manga Whisperer: Automatically Generating Transcriptions for Comics. In: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp 12967–12976, <https://doi.org/10.1109/cvpr52733.2024.01232>
- Sachdeva R, Shin G, Zisserman A (2024) Tails Tell Tales: Chapter-Wide Manga Transcriptions with Character Names. In: *Computer Vision – ACCV 2024*. Springer Nature Singapore, pp 63–80, https://doi.org/10.1007/978-981-96-0908-6_4

- Sap M, Prasettio MC, Holtzman A, et al (2017) Connotation Frames of Power and Agency in Modern Films. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp 2329–2334, <https://doi.org/10.18653/v1/d17-1247>
- Shan S, Cryan J, Wenger E, et al (2023) Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models. In: 32nd USENIX Security Symposium (USENIX Security 23). USENIX Association, pp 2187–2204, URL <https://www.usenix.org/conference/usenixsecurity23/presentation/shan>
- Shan S, Ding W, Passananti J, et al (2024) Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generative Models. In: 2024 IEEE Symposium on Security and Privacy (SP). IEEE, pp 807–825, <https://doi.org/10.1109/SP54263.2024.00207>
- Törnberg P (2023) ChatGPT-4 Outperforms Experts and Crowd Workers in Annotating Political Twitter Messages with Zero-Shot Learning. [arXiv:2304.06588](https://arxiv.org/abs/2304.06588)
- Underwood T, Bamman D, Lee S (2018) The Transformation of Gender in English-Language Fiction. *Journal of Cultural Analytics* 3(2). <https://doi.org/10.22148/16.019>
- Unser-Schutz G (2015) What text can tell us about male and female characters in shōjo- and shōnen-manga. *East Asian Journal of Popular Culture* 1(1):133–153. https://doi.org/10.1386/eapc.1.1.133_1
- Vivoli E, Bertini M, Karatzas D (2024) CoMix: A Comprehensive Benchmark for Multi-Task Comic Understanding. In: Advances in Neural Information Processing Systems 37. Neural Information Processing Systems Foundation, Inc. (NeurIPS), pp 140828–140846, <https://doi.org/10.52202/079017-4470>
- Vivoli E, Llabrés A, Souibgui MA, et al (2025) ComicsPAP: Understanding Comic Strips by Picking the Correct Panel. In: Document Analysis and Recognition – ICDAR 2025. Springer Nature Switzerland, pp 337–350, https://doi.org/10.1007/978-3-032-04614-7_19
- Wertham F (1954) *Seduction of the Innocent*. Rinehart, New York
- Wigard J, Arnold T, Tilton L (2023) Understanding Peanuts and Schulzian Symmetry: Panel Detection, Caption Detection, and Gag Panels in 17,897 Comic Strips Through Distant Viewing. *Journal of Cultural Analytics* 8(3). <https://doi.org/10.22148/001c.87560>
- Wright BW (2001) *Comic Book Nation: The Transformation of Youth Culture in America*. Johns Hopkins University Press, Baltimore

- Wu J, Tang C, Wang J, et al (2025) DiffSensei: Bridging Multi-Modal LLMs and Diffusion Models for Customized Manga Generation. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 28684–28693, <https://doi.org/10.1109/cvpr52734.2025.02671>
- Ye J, Cao L, Chen D, et al (2026) Stop Drawing Scientific Claims from LLM Social Simulations Without Robustness Audits. [arXiv:2605.18890](https://arxiv.org/abs/2605.18890)
- Ziems C, Held W, Shaikh O, et al (2024) Can Large Language Models Transform Computational Social Science? Computational Linguistics 50(1):237–291. https://doi.org/10.1162/coli_a.00502