

# Gemini 3 Pro: the frontier of vision AI

Gemini 3 Pro delivers state-of-the-art performance across document, spatial, screen and video understanding.

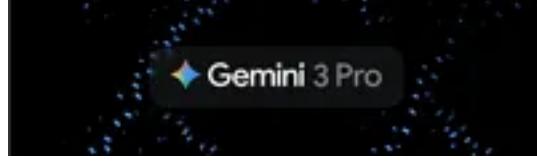


Figure 1: Image with black background and Gemini 3 Pro logo

Gemini 3 Pro represents a generational leap from simple recognition to true visual and spatial reasoning. It is our most capable multimodal model ever, delivering state-of-the-art performance across document, spatial, screen and video understanding.

This model sets new highs on vision benchmarks such as MMMU Pro and Video MMMU for complex visual reasoning, as well as use-case-specific benchmarks across document, spatial, screen and long video understanding.

Category	Benchmark	Description	Gemini 3 Pro	Gemini 2.5 Pro	Claude Opus 4.5	GPT-5.1
Visual Reasoning	MMMU Pro	Robust multimodal understanding and visual reasoning	<b>81.0%</b>	68.0%	72.0%	76.0%
	VLMsAreBiased	Visual evidence grounding versus memorized priors	<b>50.6%</b>	24.3%	32.7%	21.7%
Document	CharXiv Reasoning	Scientific chart understanding and reasoning	<b>81.4%</b>	69.6%	67.2%	69.5%
	OmniDocBench1.5*	Diverse PDF document parsing (OCR, layout, tables). Lower is better	<b>0.115</b>	0.145	0.120	0.147
Spatial	ERQA	Grounding objects and spatial concepts in the physical world	<b>70.5%</b>	56.0%	51.3%	60.0%
	Point-Bench	Fine-grained grounded pointing for spatial and affordance tasks	<b>85.5%</b>	62.7%	38.5%	41.8%
	RefSpatial	Spatial referring expressions with multi-step reasoning	<b>65.5%</b>	33.6%	19.5%	28.2%
	CV-Bench	Fundamental 2D/3D spatial understanding and relationship reasoning	<b>92.0%</b>	85.9%	83.8%	84.6%
	MindCube	Spatial mental modeling and reasoning from limited visual viewpoints	<b>77.7%</b>	57.5%	58.5%	61.7%
Screen	ScreenSpot Pro	GUI grounding in professional software environments	<b>72.7%</b>	11.4%	49.9%	3.50%
	Gui-World QA	Dynamic and sequential GUI understanding across diverse software environments	<b>68.0%</b>	42.8%	44.9%	38.7%
Video	Video-MMMU	Knowledge acquisition from educational videos	<b>87.6%</b>	83.6%	84.4%	80.4%
	Video-MME	Temporal reasoning and long-context video analysis	<b>88.4%</b>	86.9%	84.1%	86.3%
	1H-VideoQA	Time-specific query understanding in hour-long videos	<b>81.8%</b>	79.4%	52.0%	61.5%
	Perception Test	Perception and reasoning skills in real-world videos	<b>80.0%</b>	78.4%	74.1%	77.8%
	YouCook2	Procedural video understanding for cooking recipes	<b>222.7</b>	188.3	145.8	132.4
	Vatex	Multilingual video captioning and translation	<b>77.4%</b>	71.3%	60.1%	62.9%
	Motion Bench	Fine-grained video motion perception and comprehension (e.g., trajectory, repetition)	<b>70.3%</b>	66.3%	65.9%	61.1%
Education	Math Kangaroo	Multimodal mathematical reasoning with visual dependency	<b>84.4%</b>	77.4%	68.9%	79.9%
Biomedical	MedXpertQA-MM	Expert-level medical reasoning in clinical scenarios	<b>77.8%</b>	65.9%	62.2%	65.5%
	VQA-RAD	Visual question answering on radiology images	<b>81.9%</b>	71.4%	76.0%	72.2%
	MicroVQA	Multimodal reasoning benchmark for microscopy-based biological research	<b>68.8%</b>	63.5%	61.4%	61.5%

Eval@ default to MEDIA\_RESOLUTION\_HIGH (1120 tokens per image, 280 tokens per video frame). Exceptions: 1H-VideoQA uses MEDIA\_RESOLUTION\_MEDIUM (70 tokens per video frame) and ScreenSpot Pro uses MEDIA\_RESOLUTION\_ULTRA\_HIGH (2240 tokens per image).

## 1. Document understanding

Real-world documents are messy, unstructured, and difficult to parse —often filled with interleaved images, illegible handwritten text, nested tables, complex mathematical notation and non-linear layouts. Gemini 3 Pro represents a major leap forward in this domain, excelling across the entire document processing pipeline—from highly accurate Optical Character Recognition (OCR) to complex visual reasoning.

### Intelligent perception

To truly understand a document, a model must accurately detect and recognize text, tables, math formulas, figures and charts regardless of noise or format.

A fundamental capability is “derendering”—the ability to reverse-engineer a visual document back into structured code (HTML, LaTeX, Markdown) that would recreate it. As illustrated below, Gemini 3 demonstrates accurate perception across diverse modalities including converting an 18th-century merchant log into a complex table, or transforming a raw image with mathematical annotation into precise LaTeX code.



Figure 2: Input image of an old merchants handbook ledger along with an output image that clearly reconstructed transcription

Example 1: Handwritten Complex Table from 18th century Albany Merchant’s Handbook

Example 2: Reconstructing equations from an image

Example 3: Reconstructing Florence Nightingale’s original Polar Area Diagram into an interactive chart (with a toggle!)

### Sophisticated reasoning

Users can rely on Gemini 3 to perform complex, multi-step reasoning across tables and charts —even in long reports. In fact, the model notably outperforms the human baseline on the CharXiv Reasoning benchmark (80.5%).

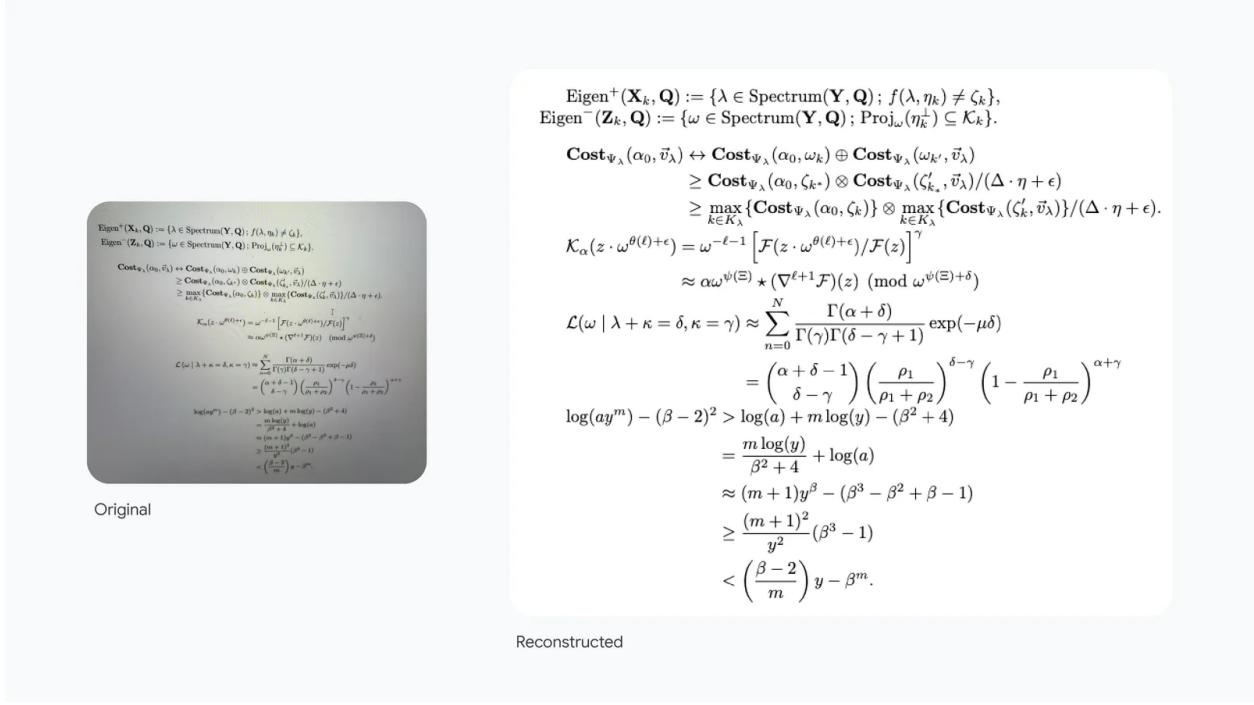


Figure 3: Input image of a scan of an equation alongside an output of the model solving the equation

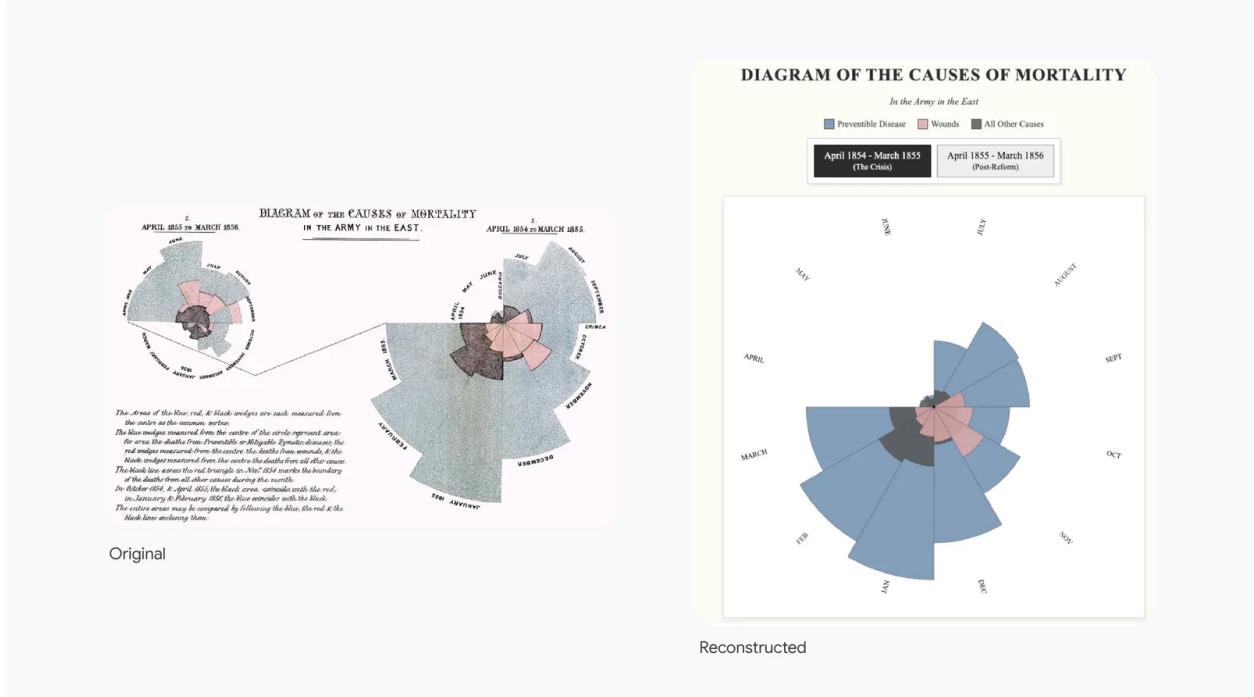


Figure 4: Image showing input of a scanned diagram into a an interactive chart

To illustrate this, imagine a user analyzing the 62-page U.S. Census Bureau "Income in the United States: 2022" report with the following prompt: "Compare the 2021–2022 percent change in the Gini index for "Money Income" versus "Post-Tax Income", and what caused the divergence in the post-tax measure, and in terms of "Money Income", does it show the lowest quintile's share rising or falling?"

Swipe through the images below to see the model's step-by-step reasoning.

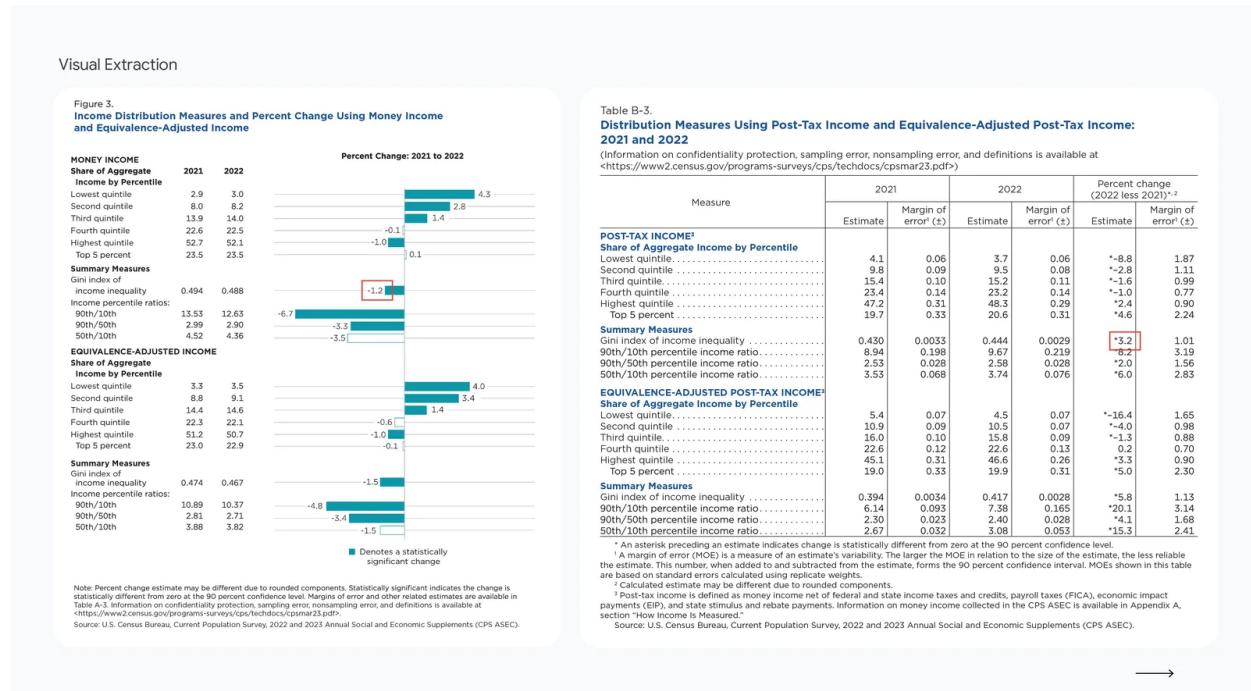


Figure 5: Pdf image highlighting the numbers -1.2 and 3.2

Visual Extraction: To answer the Gini Index Comparison question, Gemini located and cross-referenced this info in Figure 3 about "Money Income decreased by 1.2 percent" and in Table B-3 about "Post-Tax Income increased by 3.2 percent"

Causal Logic: Crucially, Gemini 3 does not stop at the numbers; it correlates this gap with the text's policy analysis, correctly identifying Lapse of ARPA Policies and the end of Stimulus Payments are the main causes.

Numerical Comparison: To compare the lowest quintile's share rising or falling, Gemini 3 looked at table A-3, and compared the number of 2.9 and 3.0, and concluded that "the share of aggregate household income held by the lowest quintile was rising."

Final Model Answer

## 2. Spatial understanding

Gemini 3 Pro is our strongest spatial understanding model so far. Combined with its strong reasoning, this enables the model to make sense of the physical world.

- Pointing capability:** Gemini 3 has the ability to point at specific locations in images by outputting pixel-precise coordinates. Sequences of 2D points can be strung together to perform complex tasks, such as estimating human poses or reflecting trajectories over time.
- Open vocabulary references:** Gemini 3 identifies objects and their intent using an open vocabulary. The most direct application is robotics: the user can ask a robot to generate spatially grounded plans like, "Given this messy table, come up with a plan on how to sort the trash." This also extends to

## Causal Logic

### APPENDIX B. POST-TAX HOUSEHOLD INCOME

The income estimates in the main sections of this report are based on the concept of money income. Money income, relative to poverty, means it does not account for tax liabilities or tax credits. Tax policies have an important effect on the total resources available to households for consumption, so an income concept that accounts for these costs and benefits is also an important measure of household wellbeing.

This appendix presents post-tax household income estimates and inequality measures for 2021 and 2022. These estimates are summarized in Tables B-1 through B-4. Post-tax income is defined as money income net of federal and state taxes, including payroll taxes (FICA), and temporary cash payments administered by tax agencies, like rebates or stimulus payments.<sup>1</sup>

Since the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) does not collect information on taxes, the Census Bureau relies on a tax calculator to simulate federal and state tax paid and credits received. Post-tax income estimates used in this appendix and the Supplemental Poverty Measure are model output from the CPS ASEC tax model. These simulations include federal and state income taxes, as well as FICA taxes, and incorporate all changes in federal and state tax laws for

calendar year 2022.<sup>2</sup> For post-tax poverty estimates, refer to the Supplemental Poverty Measure estimates in this report, "Poverty in the United States: 2022".<sup>3</sup> In 2022, several tax policies enacted by the American Rescue Plan Act (ARPA) in 2021 lapsed, including an expansion of the Earned Income Tax Credit for filers with children, the Child Tax Credit, and Child and Dependent Care Credit. In 2020 and 2021, most households also received Economic Impact Payments (EIP), but no such payments were issued in 2022.<sup>4</sup> The expiration of these policies left households with substantially less federal assistance than they received in 2021. In contrast to the direction of federal tax policy, a number of states increased assistance to households. Over 20 states issued economic rebates to their residents, and nine states expanded their state EITC and child tax credit programs.<sup>5</sup>

On net, the contraction in federal and state tax policies led to a substantial decrease in post-tax household income between 2021 and 2022. Real median post-tax household income in 2022 was 8.8 percent lower than in 2021. Refer to Table B-1 for changes in post-tax median income between 2021 and 2022 by selected demographic characteristics of the household.

Table B-2 compares median household money income estimates (without regard to poverty) to post-tax estimates by demographic characteristics of the householder in 2022. Accounting for all taxes and credits reduced median household income by 13.5 percent in 2022, compared to a 7.7 percent reduction in 2021. In 2022, all demographic groups showed statistically significant decreases in aggregate post-tax income relative to pretax income.<sup>6</sup> This contrasts with 2021, when three groups of households (female householders with no spouse present, those with a spouse aged 65 and over with no high school diploma, and those maintained by noncitizens) showed an increase in post-tax income relative to pretax income.<sup>5</sup>

Table B-3 presents post-tax inequality estimates for 2021 and 2022. In contrast to the 12.2 percent decrease in the Gini index using pretax income between 2021 and 2022 (Table A-3), the annual percentage change in the Gini index calculated by post-tax income increased 3.2 percent in 2022. In 2022, shares of aggregate post-tax income exhibited a statistically significant change at each quintile, lower, second, third, and fourth quintiles each showed a decline, while the highest quintile and the top 5 percent each showed an increase.<sup>7</sup>

Looking at the measures of equivalence-adjusted, post-tax income inequality, there was a 10 percent increase in income inequality between 2021 and 2022 as measured by the Gini index, the percentile income ratios, and the aggregate shares of money income. The share of income in the lowest income quintile increased, while the share of aggregate income of the highest quintile decreased. Inequality, as measured by the Gini index, was 9.0 percent higher when calculated using post-tax income compared to pretax income. As with pretax, equivalence-adjusted income, aggregate shares of post-tax, equivalence-adjusted income increased in the bottom four quintiles, but decreased in the highest quintile. The Gini index was also 10.6 percent lower using post-tax income instead of money income in 2022.

Comparing inequality measures using pretax money income and post-tax income in 2022 illustrates the relative impact of the tax system on income inequality. The income tax system (Table B-4). In 2022, after accounting for taxes and credits, aggregate shares of income in the three lowest income quintiles increased, while the share of aggregate income of the highest quintile decreased. Inequality, as measured by the Gini index, was 9.0 percent higher when calculated using post-tax income compared to pretax income. As with pretax, equivalence-adjusted income, aggregate shares of post-tax, equivalence-adjusted income increased in the bottom four quintiles, but decreased in the highest quintile. The Gini index was also 10.6 percent lower using post-tax income instead of money income in 2022.

### ENDNOTES

<sup>1</sup> For more information about the CPS ASEC Tax Model, refer to [www.census.gov/library/working-papers/2022/demo/SEHS-2022-008.html](https://www.census.gov/library/working-papers/2022/demo/SEHS-2022-008.html).  
<sup>2</sup> U.S. Census Bureau, "Income in the United States: 2022," [www.census.gov/library/working-papers/2023/demo/SEHS-2023-002.html](https://www.census.gov/library/working-papers/2023/demo/SEHS-2023-002.html).

<sup>3</sup> For more information about how the CPS ASEC Tax Model treats EIPs and expanded CTC and EITC, refer to [www.census.gov/library/working-papers/2023/demo/SEHS-2023-002.html](https://www.census.gov/library/working-papers/2023/demo/SEHS-2023-002.html).

<sup>4</sup> For more information about state rebate payments, refer to [www.census.gov/library/working-papers/2023/demo/SEHS-2023-002.html](https://www.census.gov/library/working-papers/2023/demo/SEHS-2023-002.html).

<sup>5</sup> CPS ASEC Tax Model treats rebates as a tax credit. The model may assign payments and credits to former tax filers who did not meet the Internal Revenue Service definition of "eligible" for stimulus payments or certain to receive stimulus payments and certain to receive stimulus payments.

<sup>6</sup> The difference between the 2021–2022 percent change for the third and fourth quartiles was not statistically significant.

42 Income in the United States: 2022

U.S. Census Bureau

U.S. Census Bureau

Income in the United States: 2022 43



Figure 6: Pdf image highlighting the ARPA policies lapsing in 2021 and the stimulus payments ending

## Numerical comparison

Table A-3.

### Income Distribution Measures Using Money Income and Equivalence-Adjusted Income: 2021 and 2022

(Information on confidentiality protection, sampling error, nonsampling error, and definitions is available at <<https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar23.pdf>>)

Measure	2021		2022		Percent change (2022 less 2021) <sup>*2</sup>	
	Estimate	Margin of error <sup>1</sup> (±)	Estimate	Margin of error <sup>1</sup> (±)	Estimate	Margin of error <sup>1</sup> (±)
<b>MONEY INCOME</b>						
<b>Share of Aggregate Income by Percentile</b>						
Lowest quintile.....	2.9	0.06	3.0	0.05	*4.3	2.50
Second quintile .....	8.0	0.09	8.2	0.08	*2.8	1.48
Third quintile.....	13.9	0.12	14.0	0.12	*1.4	1.29
Fourth quintile.....	22.6	0.17	22.5	0.17	-0.1	1.01
Highest quintile.....	52.7	0.37	52.1	0.34	*-1.0	0.93
Top 5 percent.....	23.5	0.44	23.5	0.40	0.1	2.41
<b>Summary Measures</b>						
Gini index of income inequality .....	0.494	0.0038	0.488	0.0033	*-1.2	0.98
90th/10th percentile income ratio.....	13.53	0.431	12.63	0.365	*-6.7	3.81
90th/50th percentile income ratio.....	2.99	0.034	2.90	0.046	*-3.3	1.87
50th/10th percentile income ratio.....	4.52	0.130	4.36	0.113	-3.5	3.53
<b>EQUivalence-adjusted INCOME</b>						
<b>Share of Aggregate Income by Percentile</b>						
Lowest quintile.....	3.3	0.06	3.5	0.06	*4.0	2.59
Second quintile .....	8.8	0.10	9.1	0.08	*3.4	1.41
Third quintile.....	14.4	0.12	14.6	0.11	*1.4	1.18
Fourth quintile.....	22.3	0.16	22.1	0.16	-0.6	0.92
Highest quintile.....	51.2	0.36	50.7	0.32	*-1.0	0.92
Top 5 percent.....	23.0	0.43	22.9	0.39	-0.1	2.43
<b>Summary Measures</b>						
Gini index of income inequality .....	0.474	0.0038	0.467	0.0032	*-1.5	1.03
90th/10th percentile income ratio.....	10.89	0.274	10.37	0.247	*-4.8	3.06
90th/50th percentile income ratio.....	2.81	0.034	2.71	0.037	*-3.4	1.84
50th/10th percentile income ratio.....	3.88	0.087	3.82	0.075	-1.5	2.71

\* An asterisk preceding an estimate indicates change is statistically different from zero at the 90 percent confidence level.

<sup>1</sup> A margin of error (MOE) is a measure of an estimate's variability. The larger the MOE in relation to the size of the estimate, the less reliable the estimate. This number, when added to and subtracted from the estimate, forms the 90 percent confidence interval. MOEs shown in this table are based on standard errors calculated using replicate weights.

<sup>2</sup> Calculated estimate may be different due to rounded components.

Source: U.S. Census Bureau, Current Population Survey, 2022 and 2023 Annual Social and Economic Supplements (CPS ASEC).



Figure 7: Pdf highlighting the numbers 2.9 and 3.0 for 2021 and 2022 respectively

## Final Model Answer

Based on the "Income in the United States: 2022" report, here is the comparison between the two income measures and the reasons for their divergence:

### Comparison of Gini Index Changes (2021–2022)

**Money Income (Pretax):** The Gini index decreased by 1.2 percent, moving from 0.494 in 2021 to 0.488 in 2022, representing the first annual decrease in this index since 2007.

**Post-Tax Income:** In contrast, the Gini index for post-tax income **increased by 3.2 percent**, rising from 0.430 in 2021 to 0.444 in 2022.

**Cause of Divergence in the Post-Tax Measure** The divergence in the post-tax measure was primarily caused by the expiration of federal tax policies and stimulus payments enacted in response to the COVID-19 pandemic. Specifically:

**Lapse of ARPA Policies:** Tax policies from the American Rescue Plan Act (ARPA) expired, including the expansion of the Earned Income Tax Credit (EITC) for filers without children and the full refundability of the Child Tax Credit and Child and Dependent Care Credit.

**End of Stimulus Payments:** While most households received Economic Impact Payments (EIP) in 2020 and 2021, no such payments were issued in 2022.

The expiration of these policies resulted in households receiving substantially less federal assistance than in prior years, leading to a substantial decline in real post-tax income and an increase in post-tax inequality.

**Lowest Quintile's Share of "Money Income"** In terms of **Money Income**, the share of aggregate household income held by the lowest quintile was **rising**. It increased from 2.9 percent in 2021 to 3.0 percent in 2022.

Would you like me to create a visualization comparing the income shares of different quintiles for both money income and post-tax income?

Figure 8: Final model response text

AR/XR devices, where the user can request an AI assistant to “Point to the screw according to the user manual.”

### 3. Screen understanding

Gemini 3 Pro’s spatial understanding really shines through its screen understanding of desktop and mobile OS screens. This reliability helps make computer use agents robust enough to automate repetitive tasks. UI understanding capabilities can also enable tasks like QA testing, user onboarding and UX analytics. The following computer use demo shows the model perceiving and clicking with high precision.

Task: Summarize the total revenue for each promotion type in a new sheet (Sheet2) with the promotion names as the column headers using the Pivot Table feature.

### 4. Video understanding

Gemini 3 Pro takes a massive leap forward in how AI understands video, the most complex data format we interact with. It is dense, dynamic, multimodal and rich with context.

1. **High frame rate understanding:** We have optimized the model to be much stronger at understanding fast-paced actions when sampling at >1 frames-per-second. Gemini 3 Pro can capture rapid details —vital for tasks like analyzing golf swing mechanics.

Sorry, your browser doesn’t support embedded videos, but don’t worry, you can download it and watch it with your favorite video player!

By processing video at 10 FPS—10x the default speed—Gemini 3 Pro catches every swing and shift in weight, unlocking deep insights into player mechanics.

2. **Video reasoning with “thinking” mode:** We upgraded “thinking” mode to go beyond object recognition toward true video reasoning. The model can now better trace complex cause-and-effect relationships over time. Instead of just identifying *what* is happening, it understands *why* it is happening.
3. **Turning long videos into action:** Gemini 3 Pro bridges the gap between video and code. It can extract knowledge from long-form content and immediately translate it into functioning apps or structured code.

### 5. Real-world applications

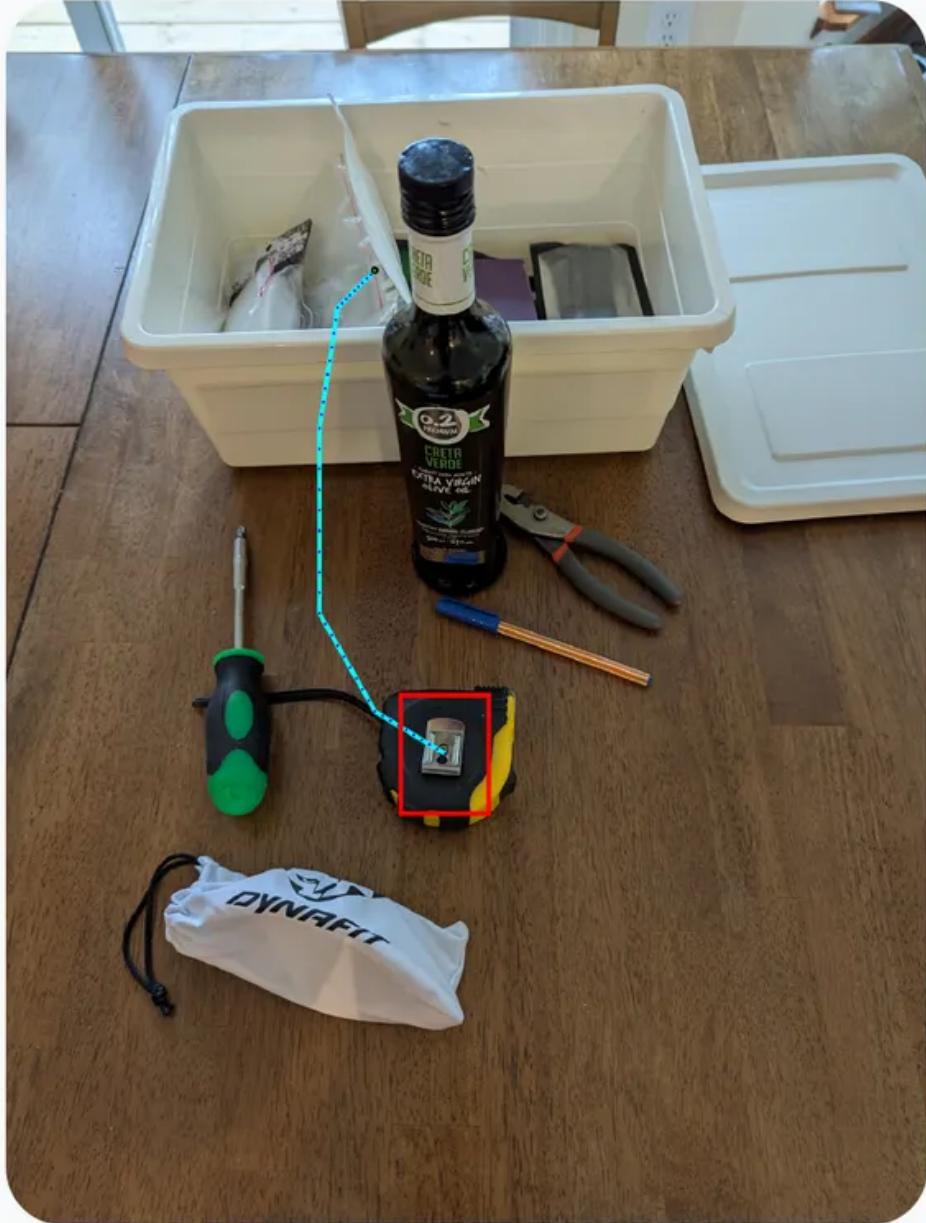
Here are a few ways we think various fields will benefit from Gemini 3’s capabilities.

#### Education

Gemini 3 Pro’s enhanced vision capabilities drive significant gains in the education field, particularly for diagram-heavy questions central to math and science. It successfully tackles the full spectrum of multimodal reasoning problems found from middle school through post-secondary curriculums. This includes visual reasoning puzzles (like Math Kangaroo) and complex chemistry and physics diagrams.

Gemini 3’s visual intelligence also powers the generative capabilities of Nano Banana Pro. By combining advanced reasoning with precise generation, the model, for example, can help users identify exactly where they went wrong in a homework problem.





## Task 1:

Generate a collision-free trajectory to put the tape measure in the bin.

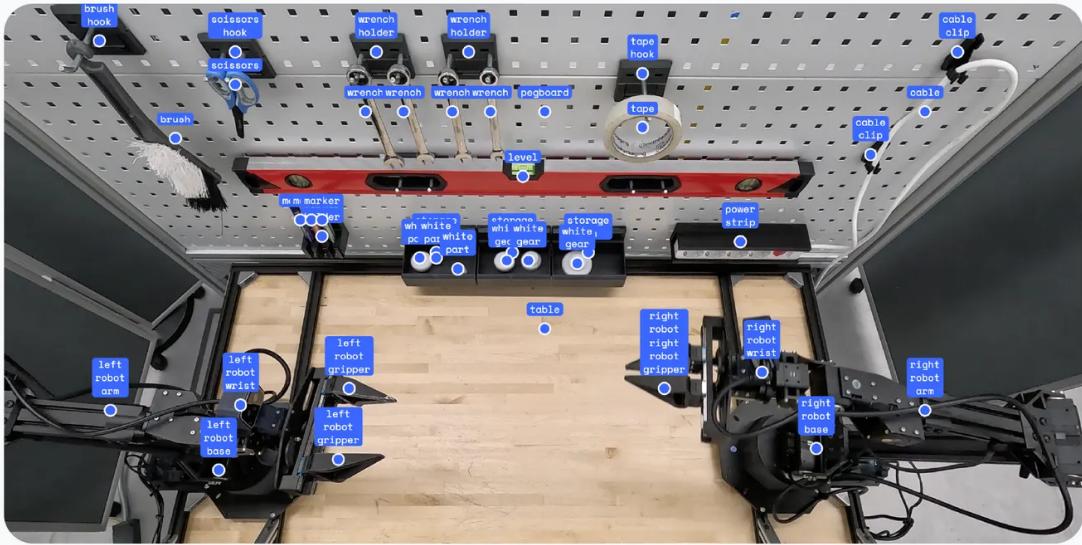
Figure 9: Image showing a cluttered box, a bottle, a screwdriver, a pouch and a measuring tape on a table. A line connects a clear path between the measuring tape and the box created by Gemini 3 Pro



## Task 2:

Point to mug, glass and bowl and where they should go.

Figure 10: A picture of a cluttered kitchen counter with open cabinets. Three lines show the trajectory between the mug, the glass and the bowl and specific<sup>10</sup> spots in the cabinet where they should go, created by Gemini 3 Pro



### Task 3:

Point to every item in this cluttered scene.

Figure 11: Picture of a circuit board with each distinct item labeled by Gemini 3 Pro

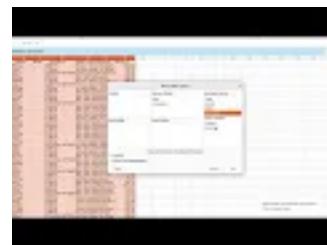


Figure 12: Video of the model interacting with an excel sheet, precisely clicking, using the cursor and typing.

Prompt: "Here is a photo of my homework attempt. Please check my steps and tell me where I went wrong. Instead of explaining in text, show me visually on my image."(Note: Student work is shown in blue; model corrections are shown in red). [ See prompt in Google AI Studio]

### Medical and biomedical imaging

Gemini 3 Pro stands as our most capable general model for medical and biomedical imagery understanding, achieving state-of-the-art performance across major public benchmarks in MedXpertQA-MM (a difficult expert-level medical reasoning exam), VQA-RAD (radiology imagery Q&A) and MicroVQA (multimodal reasoning benchmarks for microscopy based biological research).



Input image from MicroVQA - a benchmark for microscopy-based biological research

### Law and finance

Gemini 3 Pro's enhanced document understanding helps professionals in finance and law tackle highly complex workflows. Finance platforms can seamlessly analyze dense reports filled with charts and tables, while legal platforms benefit from the model's sophisticated document reasoning.

"We're impressed by Gemini 3's improvements in advanced legal reasoning, especially its ability to understand and edit contracts with complex redlines. This has been particularly valuable for our in-house customers due to the high volume and variability of the legal contracts they handle."

Harvey.ai

## 6. Media resolution control

Gemini 3 Pro improves the way it processes visual inputs by preserving the native aspect ratio of images. This drives significant quality improvements across the board.

Additionally, developers gain granular control over performance and cost via the new media\_resolution parameter. This allows you to tune visual token usage to balance fidelity against consumption:

- **High resolution:** Maximizes fidelity for tasks requiring fine detail, such as dense OCR or complex document understanding.
- **Low resolution:** Optimizes for cost and latency on simpler tasks, such as general scene recognition or long-context tasks.

For specific recommendations, refer to our Gemini 3 Documentation Guide.

## Build with Gemini 3 Pro

We are excited to see what you build with these new capabilities. To get started, check out our developer documentation or play with the model in Google AI Studio today.