MICROBLINK

Document Fraud in the Age of AI:

HOW TO MEASURE DEEPFAKE DETECTION

Executive Summary

1. AI-Generated Identity Fraud is Surging

The combination of historic data breaches and accessible generative AI tools has enabled fraudsters to mass-produce highly realistic fake IDs, posing a serious risk to businesses and security.

2. IDNet: A Breakthrough Benchmark for Deepfake Detection

The government-funded IDNet dataset is the first large-scale, standardized benchmark for evaluating identity verification systems against synthetic document fraud—including face morphs, swaps, and text tampering.

3. Face Photo Tampering Remains a Critical Threat

As generative AI tools become more advanced and accessible, fraudsters can now create hyper-realistic fake IDs with manipulated face photos that are nearly indistinguishable from real ones. These synthetic faces are designed to evade detection, making it critical for document verification systems to isolate and scrutinize the face photo itself.

4. The Importance of Achieving the Right Balance

Any business can achieve zero fraud by blocking everyone! The key is finding the right balance between user experience and security. That's why sensitivity settings matter: they allow businesses to adjust and fine tune their risk appetite appropriately.

5. Microblink's Technology Achieved a 0% False Acceptance Rate (FAR)

In evaluating its fraud detection system against the IDNet dataset, Microblink accepted zero fraudulent images—demonstrating high resilience against synthetic document threats.

Dawn of the Deepfakes

In just the past two years, data breaches have hit historic highs. The Identity Theft Resource Center reported that 2023 and 2024 saw more data compromises in the U.S. than any other years since tracking began in 2005. This creates a vicious cycle: more data breaches beget more identity fraud and the problem compounds itself.

And then when you add Generative AI to the mix, it becomes a seemingly hopeless scenario. Tools like DALL·E, Midjourney, and Stable Diffusion, originally designed for creative image generation, are increasingly being used along with stolen data from breaches to fabricate synthetic IDs with alarming realism.

This convergence of breached personal data and generative AI has opened the floodgates for scalable identity fraud. Fraudsters can now generate entire batches of synthetic identities that appear legitimate at a glance and even under digital inspection. For businesses this isn't just a security issue—it's an existential risk.

If you are unable to accurately identify deepfakes and spoofs you risk not only more fraud but a loss of consumer confidence, negative impacts to brand reputation and possibly punitive regulatory measures, such as fines.

Flurry of Fraud

The U.S. FTC's Consumer Sentinel Network, which logs reports of fraud and identity theft, received

more than 6 million

complaints from U.S. consumers in 2024.

Enter IDNet

Late in 2024, a group of researchers supported by the US Department of Homeland Security published a dataset called IDNet which represents the first large scale dataset (consisting of more than 800,000 images) of AI-generated synthetic documents. The goal of this initiative is to benchmark document verification solutions by seeing how well they can detect these AI-generated deepfakes.

At Microblink, we've been generating synthetic identity documents in our Fraud Lab for years, using AI to simulate different scenarios and build robust training datasets. To date, we've created and actively use over 280,000 synthetically generated documents to train and evaluate our models. This work helps us combat not just generative AI-based fraud, but also a wide range of digital and physical document tampering methods. So when the IDNet dataset was released, it offered a valuable opportunity to test Microblink's solution against third-party data.

In this report, we share the results of that evaluation. More importantly, anyone can test their own identity verification systems against the IDNet dataset to see how well it can detect AI-powered fraud. This is noteworthy because it marks the first time identity verification providers and enterprises alike can benchmark their fraud detection capabilities against a standardized, government-backed dataset—offering a clear, objective view into how resilient their systems are to the next generation of deepfake threats.

280,000+

number of training data images proprietary to Microblink

837,000+

number of synthetic images in the IDNet public dataset

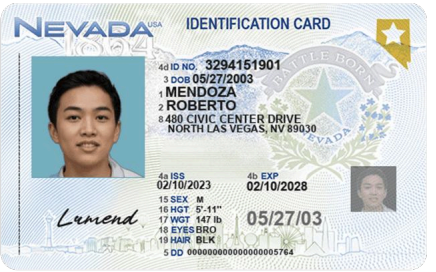
Using Data to Fight Deepfakes

The IDNet initiative signals a clear recognition: deepfakes and synthetic IDs are not merely an inconvenient problem — they're an urgent security issue for businesses across all industries.

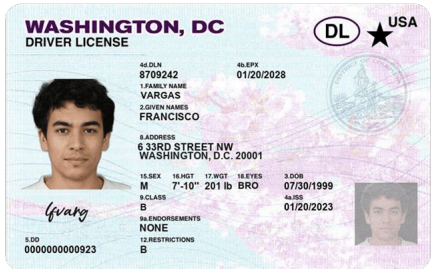
The IDNet dataset contains more than 837,000 hyper-realistic images of synthetic identity documents, totaling nearly 500 GB of data. It simulates fraud patterns that mirror what real-world attackers are doing: face morphing, portrait substitution, text field tampering, and even document-wide manipulation using generative AI.

Documents covered in the dataset include those from 10 U.S. states and European countries, featuring a variety of formats and fraud techniques. For the purposes of this exercise, Microblink tested against only the U.S. documents.

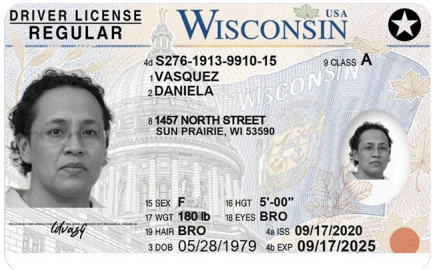
The sample lacks the face watermark and profile line pattern, has an incorrect aspect ratio, and shows cropping artifacts around the face.



The sample is missing the blended photo area, face line pattern, and bottom watermark. Its aspect ratio also differs from legitimate samples.



The image for this document should be black and white and the line security pattern should go on top of the face photo. The face photo is also overcropped.

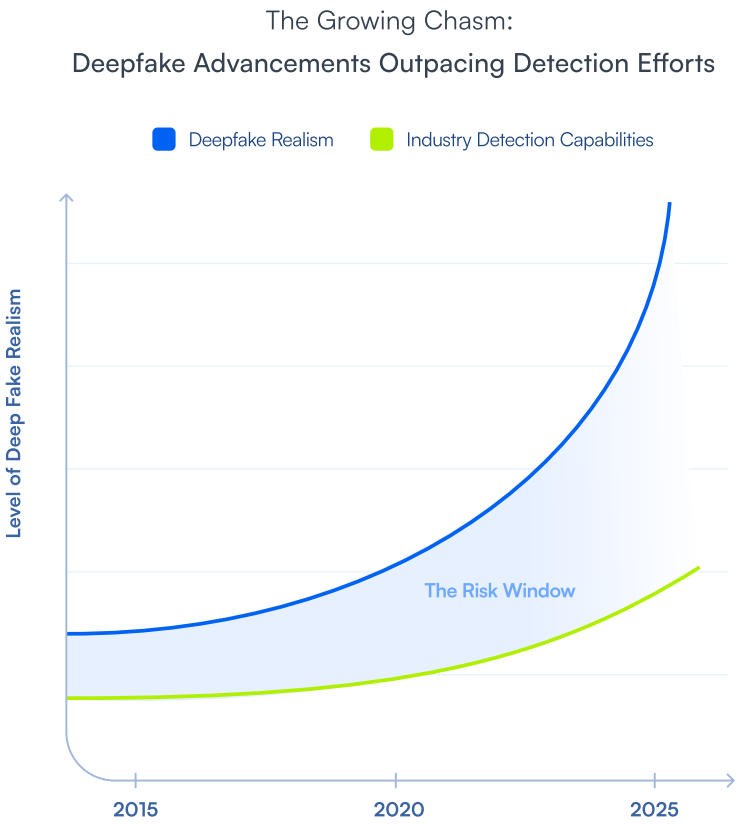


The sample lacks the face photo fade-in effect, uses grayscale instead of a warm tone, and shows cropping artifacts around the hair.

Why the Fraud and Identity Industry Needs Real Benchmarks

A likely impetus for the creation of the IDNet database is the fact that there is no official national (or international) standard for detecting deepfakes in identity verification yet. That means vendors can make big claims without a shared benchmark or a clear way to prove their effectiveness. In a market flooded with buzzwords like “AI-powered” or “deepfake detection,” it's hard for businesses to know who they can trust.

That’s why independent datasets like IDNet matter. They offer a rare opportunity to test identity verification systems against a neutral, high-stakes benchmark. A standardized testing protocol is vital for tracking algorithmic improvements across development sprints and for objective comparison as industry benchmarks emerge.



Why IDNet Is Ideal For Benchmarking Anti-Deepfake Performance

IDNet is a substantial dataset comprising a vast collection of synthetically generated identity document images, specifically engineered for training and evaluating computer vision and machine learning models for identity verification. Each image is typically accompanied by detailed metadata, including the document type, simulated issuing authority, specific fraud techniques applied, and ground truth for all data fields. This is essential for supervised learning and quantitative performance assessment.

The size of the dataset provides statistically significant volume for training deep learning models, reducing overfitting, and enabling robust performance evaluation. The labeled data is crucial for supervised training of complex deep learning architectures for tasks including document type identification, field localization, OCR, and fraud classification. The neural network architectures this data powers could include ResNet, EfficientNet for classification, U-Net for segmentation, and Transformers for sequence data in MRZs.

Importantly, Microblink has not used the raw IDNet data for model training. Our evaluation was

100% unbiased and purely performance-based

—a true measure of how our technology handles real-world fraud challenges under pressure.

Faking Fraud

Critically, IDNet is not limited to pristine, "genuine-like" document templates. It incorporates a wide array of meticulously crafted fraudulent modifications, designed to mimic real-world attack vectors and challenge specific detection algorithms.

The deliberate inclusion of diverse, high-fidelity synthetic documents and sophisticated fraudulent examples makes IDNet an invaluable resource for training, validating, and stress-testing the core algorithms of identity verification systems.

Examples of Fraud Tactics

- **Pixel-level Text Field Replacements**
  Altering critical data fields such as name, date of birth and expiration date. Fonts are carefully matched and the background texture is blended.
- **Face Morphing**
  Utilizing Generative Adversarial Networks (GANs) or similar techniques to create high-quality morphed face images that can deceive both human reviewers and basic biometric checks.
- **Strategic Portrait Substitution**
  Replacing original photos with images that may have different lighting conditions, backgrounds, or subtle digital alterations.
- **Composite Fraud Patterns**
  Combining multiple fraud techniques on a single document to test the system's ability to detect complex, multi-layered attacks.

How IDNet Data Can Help Optimize Machine Learning Model Performance

Understanding Failure Modes
Analyzing where an identity verification solution fails on IDNet's fraud samples provides critical insights into algorithmic weaknesses and potential vulnerabilities.

This allows for targeted research and development of new countermeasures.

Automated Regression Testing
IDNet can be integrated into Continuous Integration/ Continuous Deployment (CI/CD) pipelines for automated regression testing.

Any new code commit or model update can be automatically evaluated against IDNet to ensure that performance does not degrade and that previously detected fraud types are still correctly identified.

Heuristics for Zero-Day Attacks
While IDNet contains known fraud types, the process of building robust detectors for these can lead to the development of more generalized anomaly detection techniques that might have a better chance of flagging novel, unseen ("zero-day") fraud attempts.

This is essential to stay ahead of GenAI fraud whack-a-mole.

Privacy-Centric AI Development
The synthetic nature of IDNet directly addresses and mitigates significant privacy concerns and regulatory restrictions associated with using real identity documents containing PII.

This allows data scientists and ML engineers to conduct large-scale experiments, hyperparameter tuning, and architectural explorations without the ethical and legal complexities of handling vast quantities of real, sensitive user data, accelerating the R&D cycle.

Mitigating Bias and Improving Fairness
Analyzing model performance across different document types and the demographic proxies that can be inferred from photo characteristics within IDNet can help identify and mitigate potential model biases, ensuring fairer and more equitable outcomes.

Demonstrable Due Diligence for KYC/AML
Robust, documented testing against comprehensive and challenging datasets that include sophisticated fraudulent examples helps businesses demonstrate a high level of due diligence to regulators, supporting their compliance with stringent KYC and AML requirements.

IDNet in Action:
Testing the Microblink Identity Solution

Tuning for IDNet: Testing Within Real Constraints

All of the images in IDNet are tightly cropped and only show the front of the document—conditions that would typically trigger a rejection in Microblink's system due to quality and completeness standards. But to mirror the dataset's constraints and maintain fairness, we adjusted our configurations to:

- Accept front-only document submissions
- Allow tightly cropped images without immediate rejection

In real-world scenarios, ID documents are typically captured in context—often via a smartphone camera—with the surrounding environment and background visible. Microblink leverages this background data as part of its fraud detection logic. A fully cropped image with no visible context is normally flagged as suspicious or even rejected outright, as it can be a strong indicator of a fake or manipulated ID.

Additionally, our system is designed to process both the front and back of an ID, cross-checking data across both sides to detect inconsistencies or signs of tampering. Because IDNet contains only front-side images, we also disabled this part of the fraud detection stack during evaluation.

These adjustments ensured that our assessment of IDNet was fair, accurate, and reflective of the dataset's limitations —while still providing insight into how our system performs under extreme, high-risk conditions.

Not All “Genuine” Docs Are Real Anymore

IDNet includes a category labeled as bona fide documents— examples of supposedly genuine documents. However, they also were created using AI. They are meant to represent documents without tampering, but in reality they’re still artificially created nonetheless.

That’s why we treated all IDNet samples as fraudulent. It’s a tougher standard, but it better reflects the real world, where fraudsters are constantly leveling up—and even a document that “looks real” may be anything but.

Microblink’s Approach to Testing IDNet

We ran two types of evaluations:

- **End-to-end analysis**, using all of our fraud signals (not just face photo analysis)
- **Focused testing,** isolating performance on face photo tampering specifically

For both, we used standard metrics: False Acceptance Rate (FAR) and False Rejection Rate (FRR). We also looked at:

- **Image Quality Rejection Rate**
- **Microblink's Recommended Outcome Score** (our confidence signal for document authenticity)

Achieving the Right Balance

Of course, any business can completely eliminate fraud by rejecting everyone! But that is a pyrrhic victory, to say the least. It's easy to focus only on how well a solution catches fraud—but that's only half the story. The real test is balancing security with user experience. This is where the False Acceptance Rate (FAR) becomes crucial.

FAR measures how often a system incorrectly accepts a fraudulent document as legitimate.

A high FAR means more bad actors are slipping through—posing serious risks to both compliance and customer safety. While user experience matters, stopping fraud is non-negotiable.

Even a small increase in FAR can translate to significant real-world consequences.

The IDNet dataset consists entirely of AI-generated identity documents—some labeled as bona fide by the creators, but all ultimately synthetic. Because there are no true "real" documents in the dataset, calculating an actual FRR from this data isn't feasible.

We can only measure False Acceptance Rate (FAR) — how often fraudulent documents are mistakenly accepted. These are also known as false positives.

But evaluating only FAR would create an unrealistic picture. It's easy to build a system that catches all fraud—if you don't care about falsely rejecting genuine users. That's not a viable tradeoff in the real world.

To provide a more complete context, we're including the expected FRR for our system at the default sensitivity settings used in this evaluation. These settings reflect our standard production environment, balancing fraud detection with a seamless user experience. While each fraud check in our system can be tuned to be stricter or more lenient, the default configuration is designed for optimal performance across industries.

Apart from measuring FAR/FRR rates, the IDNet dataset can be used to measure model performance metrics such as:

- **F1-Score:** Harmonic mean of precision and recall, crucial for imbalanced datasets (where fraud is rare).
- **ROC AUC (Receiver Operating Characteristic Area Under Curve):** Evaluates the trade-off between true positive rate and false positive rate across different decision thresholds.
- **Precision-Recall Curves:** Particularly informative for fraud detection tasks.
- **Character Error Rate (CER) / Word Error Rate (WER):** For OCR performance.

STATE	DOCUMENT TYPE	EXPECTED FRR
Arizona	Driver's License	2.17%
California	Driver's License	1.38%
District of Columbia	Driver's License	3.11%
Nevada	Identity Card	1.56%
North Carolina	Driver's License	3.30%
Pennsylvania	Driver's License	2.76%
South Dakota	Driver's License	2.17%
Utah	Driver's License	1.46%
West Virginia	Driver's License	2.66%
Wisconsin	Driver's License	2.70%

How Did We Do?

The table below summarizes our evaluation using the “RecommendedOutcome” signal—our system’s suggested next step for each document.

Most importantly, not a single image was marked as “Accept”, meaning our False Acceptance Rate (FAR) was 0% across the tested dataset.

While the ideal result for every image is “Reject,” a portion fell into “Unprocessed,” “Undeterminable,” or “Retry” due to issues like image quality. A deeper review of these cases is outlined in the table to the right.

DOCUMENT TYPE	TOTAL IMAGES	UNPROCESSED	UNDETERMINABLE	REJECT	ACCEPT	RETRY	FAR
Arizona Driver’s License	41,852	23.08%	0.00%	72.79%	0.00%	4.13%	0.00%
California Driver’s License	41,852	0.29%	0.01%	99.70%	0.00%	0.00%	0.00%
District of Columbia Driver’s License	41,852	0.00%	0.01%	99.99%	0.00%	0.00%	0.00%
Nevada Identity Card	41,852	0.00%	0.08%	99.90%	0.00%	0.02%	0.00%
North Carolina Driver’s License	41,852	0.00%	14.73%	85.22%	0.00%	0.05%	0.00%
Pennsylvania Driver’s License	41,852	0.00%	0.00%	99.99%	0.00%	0.00%	0.00%
South Dakota Driver’s License	41,852	0.00%	1.01%	98.98%	0.00%	0.00%	0.00%
Utah Driver’s License	41,852	0.00%	0.26%	99.74%	0.00%	0.00%	0.00%
West Virginia Driver’s License	41,852	0.00%	0.00%	99.98%	0.00%	0.02%	0.00%
Wisconsin Driver’s License	41,852	23.70%	27.09%	49.21%	0.00%	0.00%	0.00%

What Leads to False Positives?

Though Microblink successfully avoided any false acceptances, the IDNet dataset does let us examine some of the issues that lead to them. These types of errors are among the most dangerous in identity verification, as they let fraudulent identities slip through undetected.

While none of these flawed samples were accepted by our platform, we determined that certain cases—such as heavy cropping or partial field visibility—might warrant a prompt for image resubmission rather than immediate rejection.

This includes images that were missing a significant part at the bottom, making them overcropped. Many others had some mandatory fields missing or misplaced.

By analyzing cases in the IDNet dataset, we can better understand what typically causes false positives in document fraud detection systems:

- **Tightly cropped or incomplete images**
  Cropped edges can remove key security features or contextual cues, making it harder for a model to detect tampering.

- **Highly realistic face swaps or morphs**
  AI-generated manipulations that preserve lighting, texture, and facial geometry can trick systems that rely on pixel-level inconsistencies.

- **Inconsistent or low-quality image capture**
  Blurry or poorly lit photos—common in mobile uploads—can obscure telltale signs of fraud.

- **Unusual but legitimate document variation**
  Genuine edge cases (e.g., old ID formats, damaged documents) may resemble fraudulent patterns, complicating classification.

- **Overly permissive configuration settings**
  Systems set to prioritize user experience (low friction) may unintentionally allow more risky documents through.

Zeroing In on Face Swaps and Photo Tampering

Detecting face swaps and photo tampering is one of the most urgent challenges in modern identity verification. As generative AI tools become more advanced and accessible, fraudsters can now create hyper-realistic fake IDs with manipulated face photos that are nearly indistinguishable from real ones. These synthetic faces are designed to evade detection, making it critical for document verification systems to isolate and scrutinize the face photo itself.

For this reason, we conducted a focused evaluation of Microblink's performance on the core task the IDNet dataset was built to measure: detecting face photo tampering. Unlike broader document fraud—which can involve layout changes, hologram manipulation, or text edits—IDNet targets only face swaps and facial image manipulations.

Focused Evaluation

To align with this intent and eliminate any bias from surrounding document features, we intentionally excluded all non-face-related signals from our analysis. That means no image quality checks, no metadata, and no processing status flags. The evaluation was based solely on the model's ability to detect when the face photo itself had been tampered with.

This test offers a clear, unbiased view of how well Microblink's AI performs against one of the most pressing—and increasingly common—fraud vectors in the industry.

STATE	DOCUMENT TYPE	NUMBER OF SAMPLES	FALSE ACCEPTANCE RATE (FAR)
Arizona	Driver's License	8,365	3.35%
California	Driver's License	8,365	0.00%
District of Columbia	Driver's License	8,365	0.00%
Nevada	Identity Card	8,365	0.38%
North Carolina	Driver's License	8,365	0.01%
Pennsylvania	Driver's License	8,365	0.01%
South Dakota	Driver's License	8,365	0.00%
Utah	Driver's License	8,365	0.06%
West Virginia	Driver's License	8,365	0.00%
Wisconsin	Driver's License	8,365	0.00%

Conclusion: Don’t Just Take Our Word For It...

This report was based on Microblink’s evaluation of the IDNet dataset using our identity fraud detection technology. But the true power of IDNet lies in its openness and accessibility—it’s a publicly available dataset, which means **any company can test their own systems against it**.

We encourage identity verification providers, researchers, and practitioners to run their own benchmarks using IDNet. By doing so, you can:

- **Stress-test your models** against a wide range of realistic, synthetic fraud scenarios
- **Compare performance** transparently using a shared reference point
- **Spot weaknesses** in your system’s sensitivity to deepfakes, face swaps, and document tampering
- **Push the industry forward** by building on a common foundation

In the contemporary digital ecosystem, where sophisticated identity fraud represents a persistent and escalating threat, the empirical validation and continuous optimization of identity verification solutions are paramount. Benchmarking against a shared dataset like IDNet not only improves internal performance—it creates accountability and raises the bar for the entire industry. And if you want to test Microblink’s technology against IDNet, please contact us today.

The IDNet dataset is available at https://arxiv.org/pdf/2408.01690