Making Millions of PDFs Accessible – The PDFix Pipeline Approach
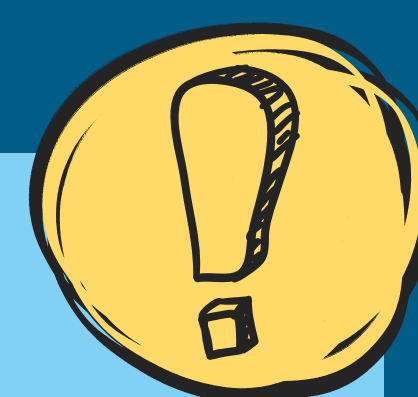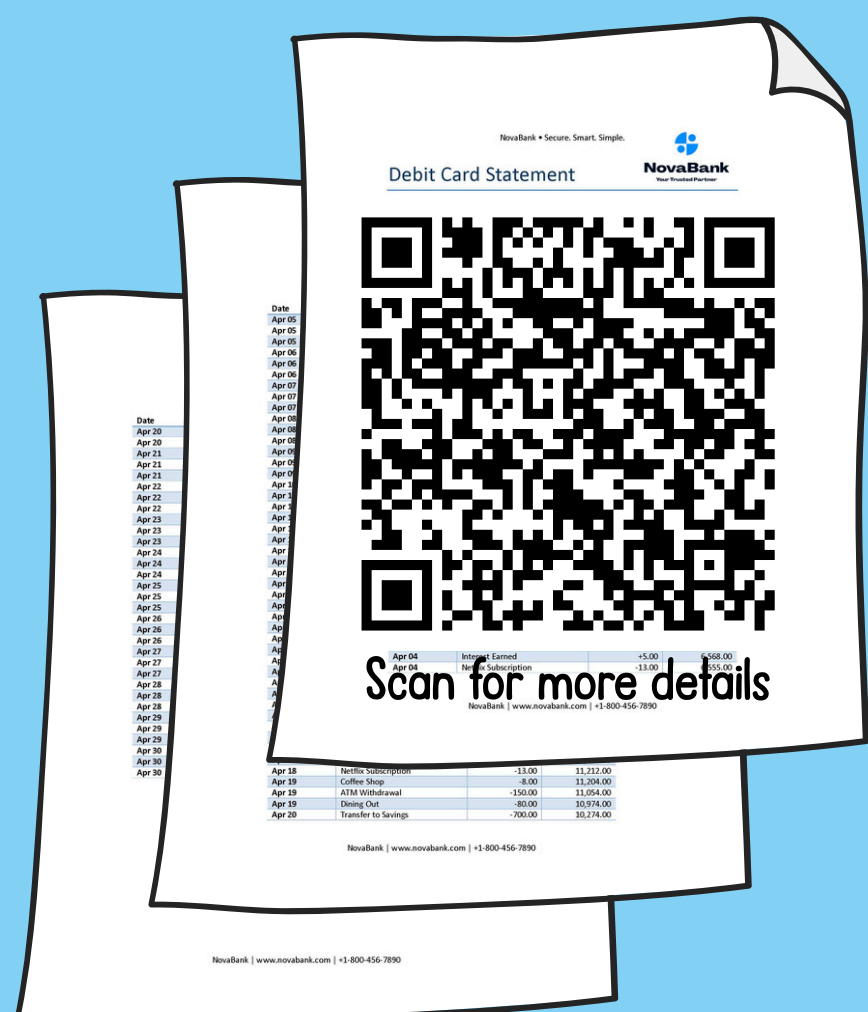
Use Case

Auto-Tag Thousands of Bank Statements

Enterprises generate millions of financial PDFs — from **bank statements** to **invoices** — yet most remain improperly tagged and inaccessible.

- Customers struggle to read statements with screen readers
- Accessibility gaps block equal access to financial information
- Organizations face compliance risks under PDF/UA and WCAG

The good news: Accessibility doesn't have to be slow or complicated. With **AI-powered, template-driven auto-tagging in PDFix**, even complex financial documents can be transformed into **fully accessible, screen-reader-friendly PDFs** — automatically, at scale, and without deep accessibility expertise.

Use Case

Auto-Tag Thousands of Invoices

Organizations process **millions of invoices** every year, but most PD[...] are delivered without proper accessibility tagging.

- Manual invoice review slows down finance teams and auditors
- Missing tags break reading order and table navigation
- Enterprises risk penalties for inaccessible financial reporting

The solution: With **AI-powered, template-driven auto-tagging in PDFix**, invoice PDFs can be processed in bulk and converted into **fully accessible, compliant documents**.

⚡ Scales across large financial workflows with minimal manual effort
⚡ Delivers faster, more accurate, and compliant invoicing at enterprise scale

PDFix Pipeline

The **PDFix Pipeline** combines the strengths of **AI** and **rule-based templates** to transform large volumes of unstructured PDFs into **accessible, compliant documents.**

1 Layout Template as the Foundation
- Defines how to detect headings, paragraphs, tables, and other structures
- Templates can be written manually or generated with AI assistance
- Ensures consistent layout recognition across millions of similar PDFs

2 AI and PDFix Actions for Enrichment
- External AI models add advanced accessibility features:
  - Alt-text generation for images
  - MathML extraction for equations
  - Table summaries for screen reader usability
  - Language detection for multilingual documents
- Internal PDFix functions handle tagging, reading order, and compliance validation

3 Accessible PDF Output
The result is a fully tagged, PDF/UA-compliant document that scales across enterprise workflows — making millions of PDFs accessible by design.

PDFix Layout Template

The **PDFix Layout Template** defines the structural logic that drives how a PDF is recognized and tagged. Instead of relying only on machine learning, it applies rules and anchors to detect headings, paragraphs, tables, figures, and captions.

- Can be written manually for precision, or generated with AI for speed
- Ensures consistent recognition across **millions of similar PDFs** — ideal for repeated formats like invoices, statements, catalogs, and reports
- Establishes the f**oundation for compliance**, guaranteeing correct reading order, semantic roles, and proper tagging

Key Benefits
- **Repeatable accuracy** → one template can be reused across large PDF collections
- **Hybrid approach** → combines AI proposals with rule-based precision
- **Compliance by design** → ensures output aligns with PDF/UA and WCAG requirements

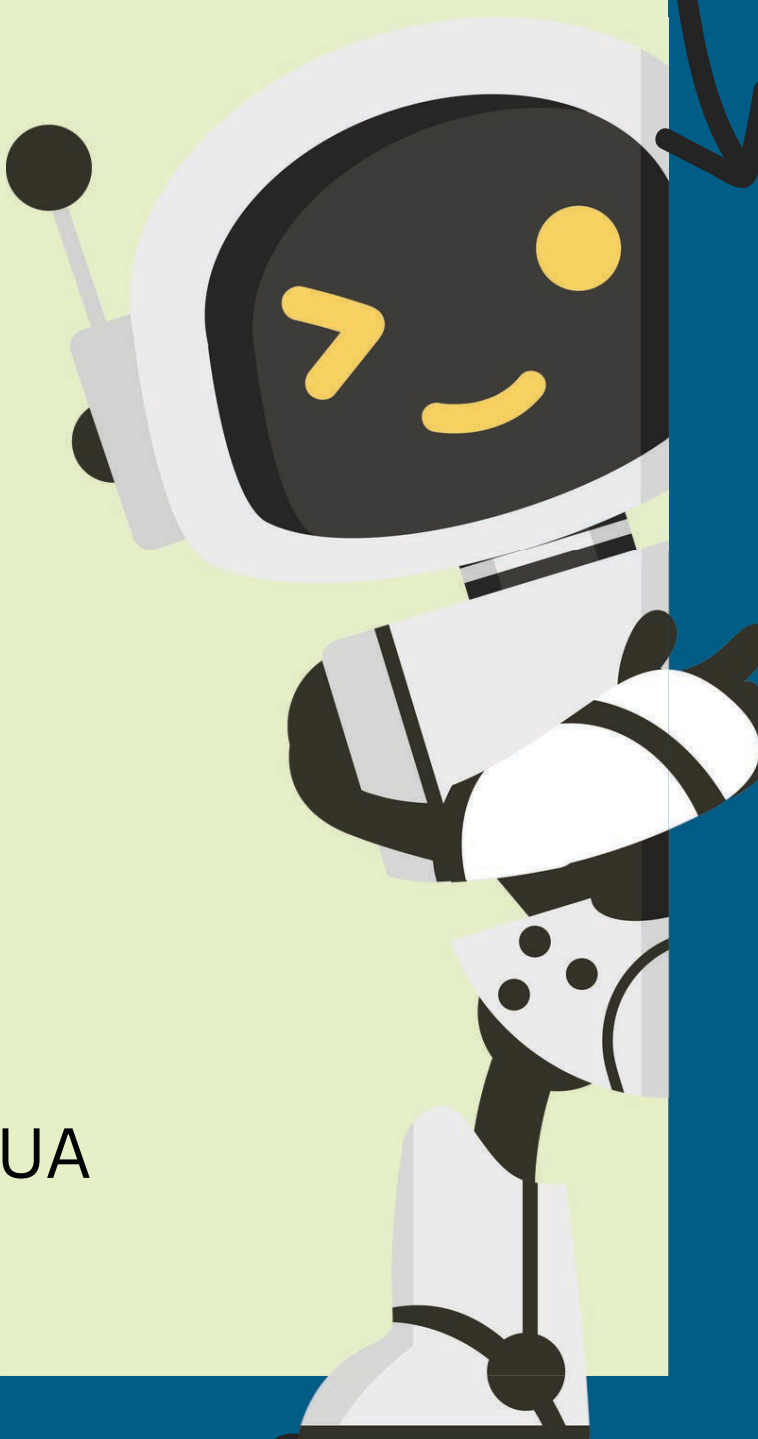### PDFix Actions

The **PDFix Desktop and SDK** can be extended through the **Actions Marketplace**, offering a growing catalog of **AI-powered external actions** — available via Docker containers or local command-line modules.

These seamlessly integrate into your workflow, enabling advanced accessibility enhancements such as:

- Auto-tagging and Layout Template detection
- Alt-text generation for images
- MathML extraction from equations
- Table summaries for screen reader clarity
- Document language detection
- Additional automation such as OCR, HTML-to-PDF/UA conversion, and validation reports
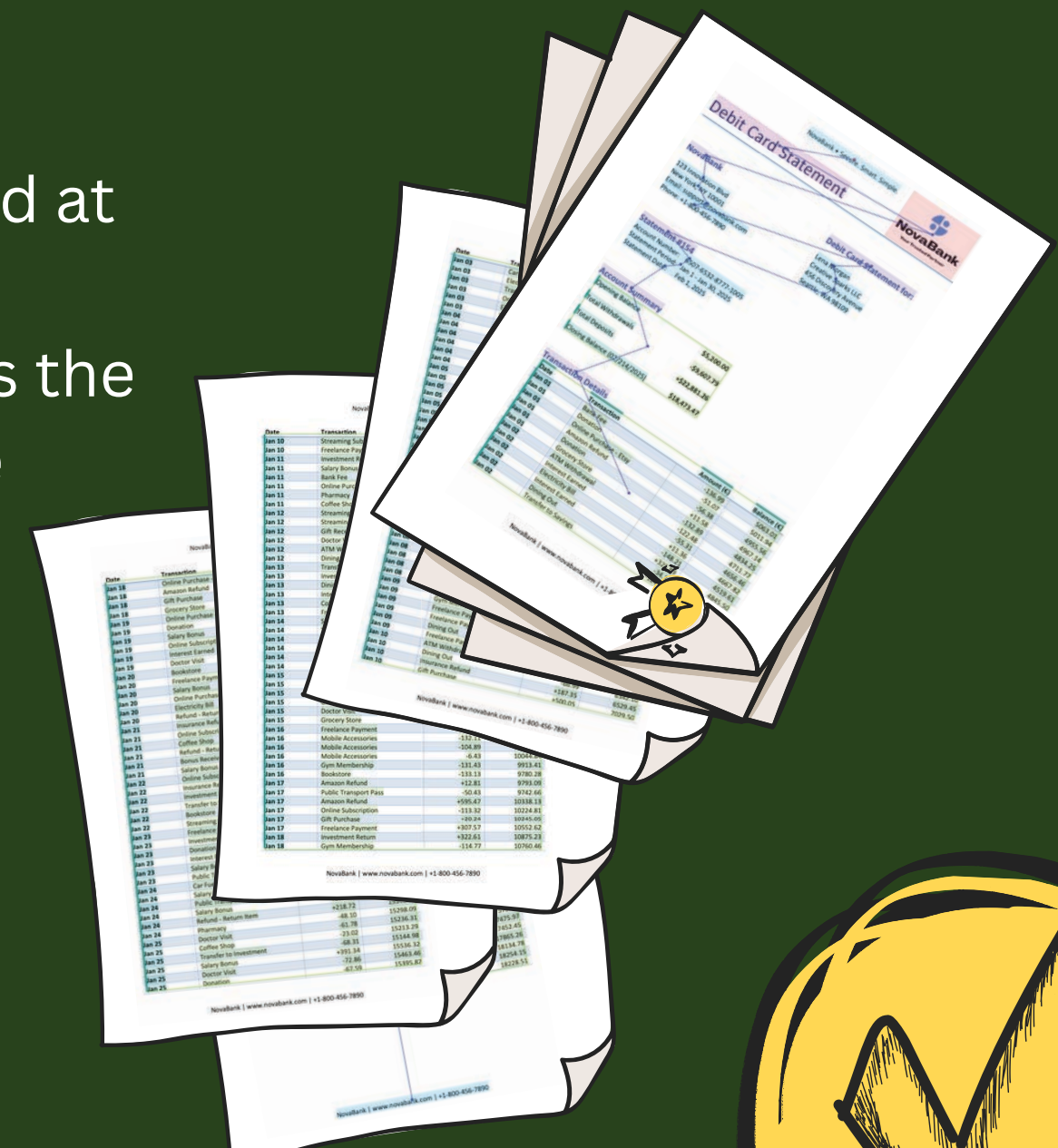
Quality and Validation

With the structure defined by a **Layout Template**, PDFix performs automated validation against accessibility standards.

- Ensures full compliance with **PDF/UA and WCAG**
- Provides clear, actionable reports so accessibility can be guaranteed at scale
- Because the layout template drives recognition, validation confirms the document is not just partially improved — but **truly fully accessible**

One pipeline, any platform, every PDF

The **PDFix Pipeline** is a **universal, multiplatform solution** that runs seamlessly on any system — delivering accessible PDFs everywhere they're needed

---

PDFix SDK
PDFix Desktop Lite
PDFix Desktop Pro
PDFix Marketplace
PDFix Web App
PDFix Pipeline
PDFix Actions

AI

www.pdfix.net

"One PDF is easy. One million is impossible. PDFix Pipeline makes it possible."

Scan for more details