

Using Large Language Models to Measure U.S. Retirement Plan Design at Scale: Methods and Evidence

Taha Choukhmane¹ John M. Dedyo² Cormac O'Dea³ Lawrence D.W. Schmidt¹

¹MIT Sloan School of Management

²Harvard University

³Yale University

NBER: AI and Economic Measurement
May 7, 2026

Why Retirement Plan Design Matters

- ~**100M Americans** save in employer sponsored DC plans.
 - Plan design—match formulas, vesting, auto-enrollment—shapes retirement wealth.
 - But comprehensive data on plan design across the universe of plans **doesn't exist**.
- Existing sources:
 - Administrative (e.g. record keeper) data: non-representation, proprietary
 - Hand-collected samples (small, often non-representative if going for scale).

The Data Problem: Needle in a Haystack

- All DC plan sponsors file annual **Form 5500** with the DOL.
 - Universe: 1.37M plans, 10.9M plan-years.
 - Each filing: \sim 100-page PDF; plan design buried in 3–5 pages of unstructured text.
- A large, but messy data set:
 - Choukhmane, Goodman, & O'Dea (2025) hand-coded \sim 6,200 plans with **20 RAs**.
 - Scaling by hand to the full universe: infeasible.

Method: Two-Stage LLM Pipeline

- **Stage 1—PDF** → **text**: OCR, regex-locate plan-description section, save 4 pages.
- **Stage 2—text** → **table** (per feature):
 - ① Fine-tune *snippet extractors* (Llama 3.1 8B) on RA-coded snippets.
 - ② Fine-tune *tabulators* using **retrieval-augmented generation**: top-5 most semantically similar hand-coded examples become in-context examples.
 - ③ **Double-code**: train two independent pipelines on disjoint partitions; accept only when both agree.

Final dataset: 146,704 plans, 21 years. **1.15M plan-years** ($\approx 20\times$ hand-collected).

Stage 1: PDF → Text

NOTES TO THE FINANCIAL STATEMENTS

1. Description of the Plan

The following description of the Best Buy Retirement Savings Plan (the "Plan") is provided for general information purposes only. Participants should refer to the Plan agreement for a more complete description of the Plan's provisions.

General - The Plan is a profit-sharing plan with a "cash or deferred" salary reduction savings arrangement intended to qualify under Internal Revenue Code (the "Code") § 401(k). Eligible employees of Best Buy Co., Inc. ("Best Buy") and subsidiaries (the "Company") may participate after reaching the age of 18. No minimum period of service is required.

The Benefits Committee ("Plan administrator") is appointed by a committee of the Board of Directors of the Company and has been delegated the Company's fiduciary and/or administrative responsibilities under the provisions of the Employee Retirement Income Security Act of 1974 ("ERISA") and the Plan. Voya Financial serves as the Plan recordkeeper and the Plan's trustee. The Plan is subject to the provisions of ERISA.

On January 1, 2024, the Current Health, Inc. 401(k) Plan was legally merged with the Best Buy Retirement Savings Plan and employees of Current Health, Inc. began participating in the Plan. During the year ended December 31, 2024, the Current Health, Inc. 401(k) Plan net assets available for benefits totaling \$3,377,825 were transferred to the Plan.

Contributions - Each year, participants may contribute up to 50% of their annual compensation through pre-tax contributions, after-tax Roth contributions or a combination of the two contribution types as defined by the Plan, subject to the Code limitations. Participants who have attained age 50 before the end of the Plan year are eligible to make catch-up contributions. Participants may also contribute amounts representing distributions from other qualified defined benefit or defined contribution plans. After one year of service with the Company, the Company will match 100% of the participant's eligible contributions that do not exceed 3% of compensation, plus 50% of eligible contributions that exceed 3% but do not exceed 5% of compensation.

Participant Accounts - Individual accounts are maintained for each Plan participant. Each participant's account is credited with the participant's contribution, the Company's matching contribution, as well as allocations of Plan earnings and losses. Participants' accounts are also charged with an administrative expense to cover expenses paid by the Plan. Charges are a flat monthly rate plus any service fees based on specific participant transactions, as defined in the Plan agreement. The benefit to which a participant is entitled to is the benefit that can be provided from the participant's vested account.

Investments - Participants direct the investment of their contributions and the Company's matching contributions into various investment options offered by the Plan, including the Best Buy Co., Inc. stock fund, registered investment companies, pooled funds and a stable value fund.

Effective September 19, 2024, the Best Buy Co., Inc. Stock Fund was converted to an employee stock ownership plan ("ESOP") and remains designated as the Best Buy Co., Inc. Stock Fund. With respect to the Best Buy Co., Inc. Stock Fund, common stock dividends paid to the ESOP will be credited to the participant based on their election. The participant may elect either to receive distribution of the dividend quarterly or reinvest the dividend in the Best Buy Co., Inc. Stock Fund. This election right applies to Best Buy Co., Inc. Stock Fund common stock dividends with a record date on or after September 19, 2024. Absent an affirmative election, a participant is deemed to have elected dividend reinvestment. A participant's election (or deemed election) is irrevocable until affirmatively changed by the participant.

There were no other changes made to the investment options of the Plan during the year ended December 31, 2024.

Vesting - Participants are immediately vested in their contributions, plus actual earnings thereon. Effective January 1, 2007, the Plan agreement was amended to adopt a safe harbor matching contribution provision intended to satisfy Section 401(k)(12)(B) of the Code. This provision provides that the participant's account balances holding such safe harbor matching contributions will be immediately 100% vested.

Notes Receivable from Participants - Employees hired on or after June 1, 2014, may not borrow from their fund accounts, and effective January 1, 2015, no participant may request a new loan under the Plan. Prior to April 1, 2014, participants could borrow from their fund accounts a minimum of \$1,000 up to a maximum equal to the lesser of \$50,000 or 50% of their vested account balance. The loans are secured by the balance in the participant's account and bear interest at the rate of the prime interest rate plus one percentage point on the first business day of the month in which the loan was processed. Loans require repayment within five years from the loan date, unless the loan was for the purchase of the participant's primary residence, in which case the repayment term is up to 15 years. Principal and interest is paid ratably through bi-weekly payroll deductions.

During the period between April 20, 2020, through September 20, 2020, the Company adopted a provision of the Coronavirus Aid, Relief and Economic Security Act (the "CARES Act") that allowed a qualified participant to request a loan from the Plan for a minimum of \$1,000 and up to a maximum of 100% of their vested balance or \$100,000 to be repaid over a period of five years.

At December 31, 2024, notes receivable from participants matured through October 19, 2029, with interest rates ranging from 4.25% to 9.00%.

Stage 1: PDF → Text

Contributions - Each year, participants may contribute up to 50% of their annual compensation through pre-tax contributions, after-tax Roth contributions or a combination of the two contribution types as defined by the Plan, subject to the Code limitations. Participants who have attained age 50 before the end of the Plan year are eligible to make catch-up contributions. Participants may also contribute amounts representing distributions from other qualified defined benefit or defined contribution plans. After one year of service with the Company, the Company will match 100% of the participant's eligible contributions that do not exceed 3% of compensation, plus 50% of eligible contributions that exceed 3% but do not exceed 5% of compensation.

Stage 2: Text \rightarrow Table

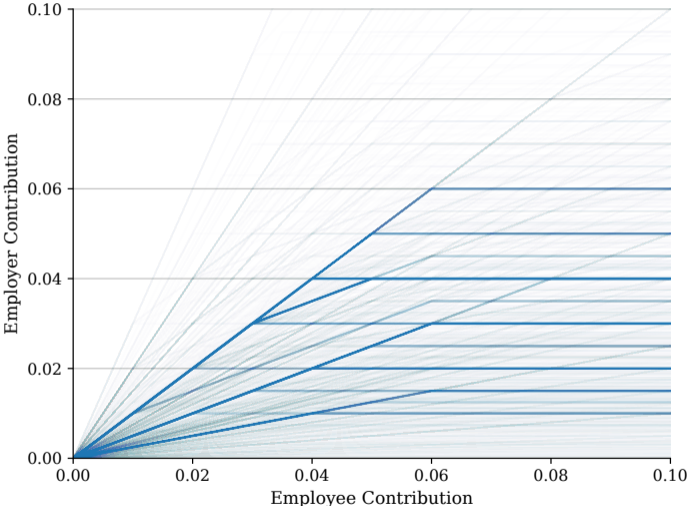
Match Rate 1	Cap 1	Match Rate 2	Cap 2	Match Rate 3	Cap 3
1.00	0.03	0.50	0.05	NA	NA

Data Quality

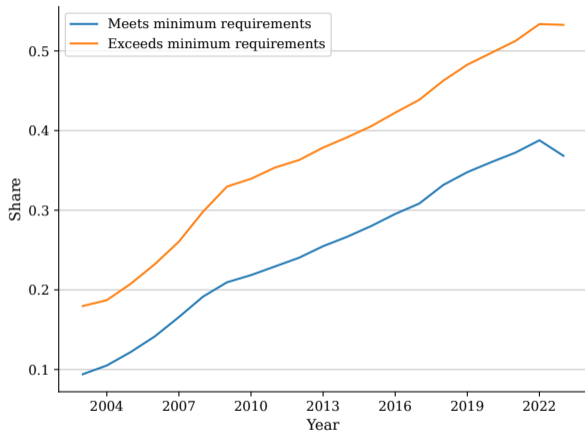
- Trained humans to blindly code ~ 450 randomly selected plans. **Perfect human-LLM agreement on 95%** of plans.

Column	N	Agree	Agreement Rate
Match Formula	439	427	97.3%
<i>Given simple match formula:</i>			
Match Rate 1	313	307	98.1%
Cap 1	313	306	97.8%
Match Rate 2	313	311	99.4%
Cap 2	313	310	99.0%
Match Rate 3	313	313	100.0%
Cap 3	313	313	100.0%

We Can See The Population



The Rise of Safe Harbor

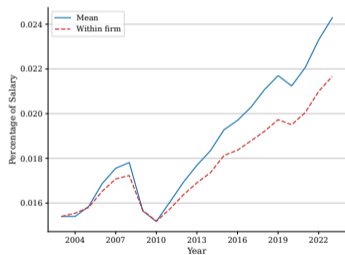


- Safe Harbor plans are the *stickiest*: 14% change match formula after 10 years vs. 32% otherwise.

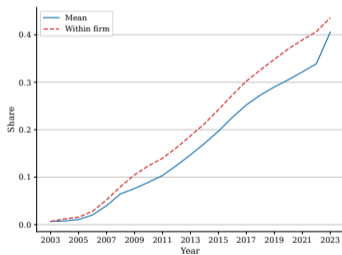
Federal regulation is reshaping plan design.

What Else This Dataset Enables

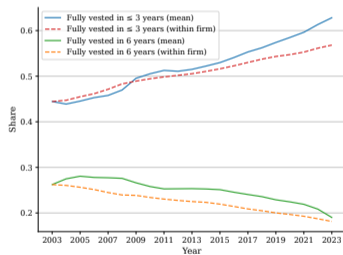
- Beyond Safe Harbor, the panel documents:
 - Match generosity rises 60% since 2003.
 - Auto-enrollment expands from $\sim 0\%$ \rightarrow 40% of plans.
 - Vesting accelerates: 45% \rightarrow 65% vest in ≤ 3 years.



(a) Maximum employer match.



(b) Auto-enrollment offered.



(c) Vesting schedule adoption.

Dataset Features

- Covers a large, growing, often unmeasured, feature of compensation
- 1000s of firm-level quasi-experiments can be used to understand how incentives shape savings behavior

LLMs make population-scale analysis possible