

AI Exposure Across Industries

A Reproducible Measurement Framework

Mario Diez Martínez

Methodological appendix · July 2026

Abstract

This appendix documents a reproducible pipeline for measuring artificial-intelligence exposure across US industries. It constructs two complementary measures: a direct industry score, AIIE, and a bottom-up occupational measure, ExpWB, that weights occupational AI exposure by wage-bill shares. Both are mapped through NAICS and SIC classifications into Fama–French 49 industry portfolios, aggregated using equal or BEA value-added weights, and standardized across covered portfolios. The purpose is measurement, not causal inference: exposure records the task content of an industry, not AI adoption, productivity, displacement, or realized returns.

1 Measurement question

Calling an industry “AI exposed” hides several choices. Exposure can be measured directly at the industry level or reconstructed from the occupations employed by that industry. Occupational scores can be weighted by headcount, wages, or both. Industry classifications also differ across the source data and the outcome datasets used later.

The pipeline makes each choice explicit and produces auditable variants. Its final unit is a Fama–French 49 portfolio because that classification can be joined to industry return data, but the intermediate NAICS-4 measures are retained for other labour- and real-economy applications.

2 Data and vintages

Input	Vintage	Role
Felten–Raj–Seamans AIOE/AIIE	2021 and 2023	Occupational and direct industry AI-exposure scores; the 2023 vintage focuses on language-model capabilities.
BLS OEWS	May 2021	Employment and mean annual wages by occupation-industry cell.
BEA GDP by Industry	2021	Value added used for alternative portfolio aggregation.
Census SIC–NAICS concordance	SIC 1987 / NAICS 2002	Classification bridge between source industries and portfolio definitions.
Ken French Data Library	Current definitions	SIC ranges defining the 49 industry portfolios.

Table 1: Inputs used in the public construction pipeline.

3 Classification chain

The mapping proceeds in two explicit steps:

1. Reverse the Census SIC-to-NAICS concordance to associate each NAICS-4 industry with its candidate SIC codes.
2. Assign each SIC code to a Fama–French portfolio and give the NAICS-4 industry the modal portfolio among its candidate SIC assignments.

This produces a many-to-one NAICS-4-to-FF49 crosswalk. Forty-three of the 49 portfolios receive at least one valid NAICS-4 match; six remain missing and are not imputed. The public repository retains the crosswalk and aggregate mapping diagnostics.

4 Direct industry exposure: AIIE

Felten, Raj, and Seamans provide AI Industry Exposure scores at mixed NAICS detail. Codes are truncated to four digits, duplicate subindustry observations are averaged within NAICS-4, and the resulting industry measure is mapped to FF49.

Let $x_{i,t}^{AIIE}$ be the direct exposure of NAICS-4 industry i in vintage $t \in \{2021, 2023\}$. No employment or wage weights enter this first-stage score. It is therefore a direct reading of the task content assigned to the industry in the source dataset.

5 Occupational exposure weighted by wage bill: ExpWB

The second measure begins with occupational exposure. For occupation o in industry i , the wage bill is

$$WB_{io} = L_{io} w_{io}, \quad (1)$$

where L_{io} is OEWS employment and w_{io} is mean annual pay. The occupation’s share of the industry wage bill is

$$\omega_{io} = \frac{WB_{io}}{\sum_{o \in i} WB_{io}}. \quad (2)$$

ExpWB is then

$$ExpWB_{i,t} = \sum_{o \in i} AIOE_{o,t} \omega_{io} = \frac{\sum_{o \in i} AIOE_{o,t} L_{io} w_{io}}{\sum_{o \in i} L_{io} w_{io}}. \quad (3)$$

The index is dimensionless. Wage-bill weighting gives more influence to occupations that account for more employment and pay inside the industry; it does not turn exposure into a dollar loss or a forecast of automation.

6 Portfolio aggregation

For any NAICS-level measure x_i , the equal-weighted portfolio score is

$$x_p^{EW} = \frac{1}{N_p} \sum_{i \in p} x_i, \quad (4)$$

where N_p is the number of mapped NAICS-4 industries. The alternative value-added-weighted score is

$$x_p^{VW} = \frac{\sum_{i \in p} x_i VA_i}{\sum_{i \in p} VA_i}, \quad (5)$$

where VA_i is 2021 BEA value added. If value added is unavailable for all industries in a portfolio, the implementation records the fallback to equal weighting rather than silently dropping the portfolio.

Every variant is standardized across portfolios with valid exposure:

$$z_p = \frac{x_p - \bar{x}}{s_x}. \quad (6)$$

Variable names preserve the construction: measure, vintage, aggregation rule, and an optional `_z` suffix. For example, `expwb_23_vw_z` is the 2023 occupational measure aggregated with value-added weights and standardized across portfolios.

7 Cross-measure validation

The high correlation is a useful internal validation because AIIE and ExpWB reach the industry from different directions. It is not mechanical equality. Divergent portfolios remain informative: they reveal places where a direct industry classification and the occupation mix put different weight on exposure.

8 Reproducibility and boundaries

The public release includes construction code, source adapters, the NAICS-to-FF49 mapping, NAICS-level aggregate measures, portfolio-level outputs, and figures. Reproduction requires downloading the documented public source files and supplying a personal BEA API key outside version control.

The following boundaries are intentional:

- Exposure is not observed adoption, productivity, substitution, employment loss, or a causal treatment.
- Six FF49 portfolios are excluded rather than assigned through an unsupported match.
- Modal SIC assignment resolves a many-to-many concordance but can hide heterogeneous subindustries.

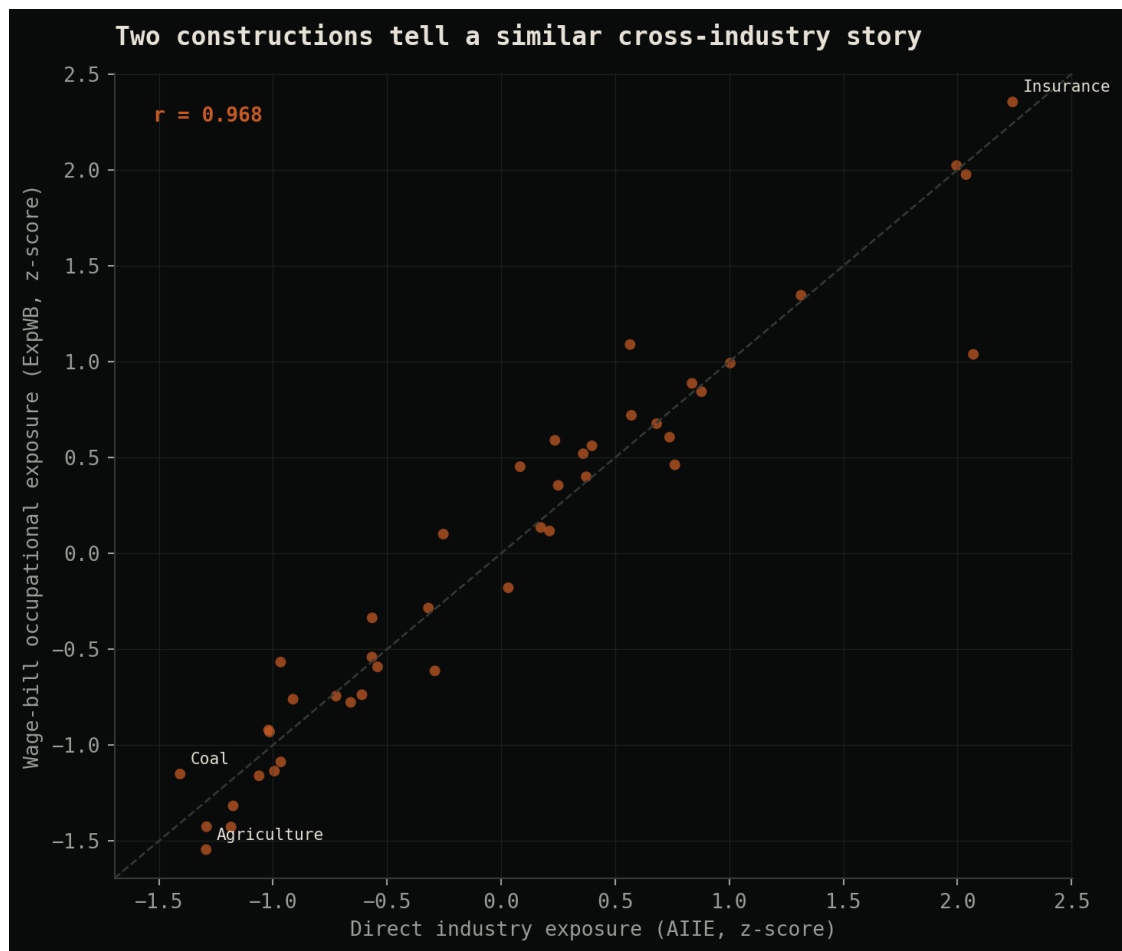


Figure 1: Direct industry exposure and wage-bill-weighted occupational exposure across the 43 covered Fama–French portfolios. The 2023 equal-weighted variants have correlation $r = 0.968$.

- OEWS employment and wage weights are fixed at the May 2021 vintage; occupational composition can change over time.
- Equal and value-added weighting answer different questions and are published side by side.
- Downstream return analysis remains experimental and is not part of the headline measurement claim.

References

- Felten, E., Raj, M., and Seamans, R. (2021). Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses. *Strategic Management Journal*, 42(12), 2195–2217.
- Felten, E., Raj, M., and Seamans, R. (2023). How Will Language Modelers Like ChatGPT Affect Occupations and Industries? Working paper.
- US Bureau of Labor Statistics. Occupational Employment and Wage Statistics, May 2021.
- US Bureau of Economic Analysis. GDP by Industry, 2021 value-added tables.
- US Census Bureau. 1987 SIC to 2002 NAICS concordance.
- Kenneth R. French Data Library. Definitions of 49 Industry Portfolios.

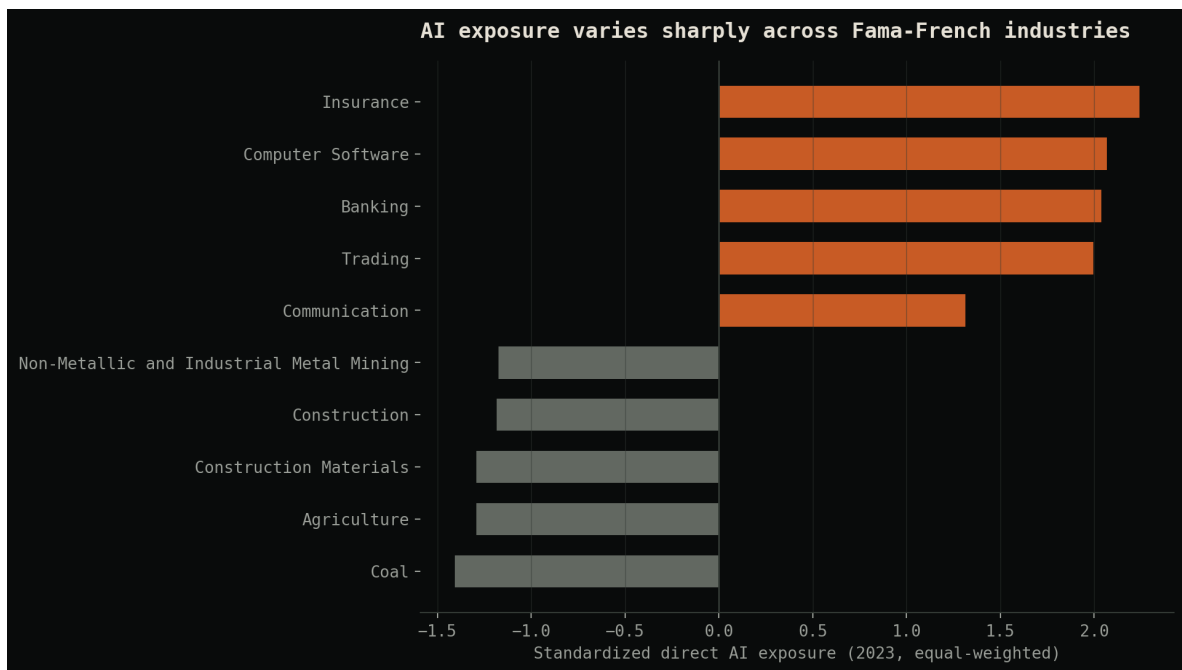


Figure 2: Selected high- and low-exposure portfolios using standardized 2023 equal-weighted AIE. The chart shows relative cross-industry exposure, not adoption or economic impact.