Mehr Aufwand, mehr Fremdvergabe, keine Steuergröße

Entwicklungsaufwand in der deutschen Automobilindustrie: Befunde, mögliche Wurzeln und Lösungswege nach dem Lebensfähigkeitsmodell

Thomas Arends · Rechercheprotokoll · Stand 27.11.2026 ·

Auf einen Blick

Befund	Zahl	Stufe
Entwicklungsaufwand je Fahrzeug steigt real, der Absatz sinkt	VW 2016–2025: F&E je Fahrzeug real +27 %, Absatz −13 %	A
Je Modell steigt er noch stärker	VW: F&E je Modell real 42,8 → 106,1 Mio. €, Faktor 2,5 (Modellabgrenzung ungeprüft)	A / E
Mehr wird außer Haus vergeben	Fahrzeugbau (WZ 29) 2016–2023: externe F&E +86 %, interne +39 %; Anteil extern 30,9 % → 37,6 %	C
Je internem Kopf weniger Budget	Interne F&E je Vollzeitäquivalent real −18,6 %	C
Verwaltung wächst am schnellsten	VW Verwaltungskosten je Fahrzeug real +53 %, Quote 3,4 % → 3,9 % vom Umsatz	A
Anforderungen sind Prosa	Bosch: 61 % der Anforderungsdefekte unkorrekt oder unvollständig; MAN: 46 % der Review-Befunde betreffen Betriebsmodi	B
Der Risikobericht zeigt es nicht	VW stuft das Entwicklungsrisiko 2020–2024 fünfmal identisch ein, während Mehrkosten in Milliardenhöhe bekannt sind	A / E
Die Vergütung misst es nicht	Vorstandsbonus je zur Hälfte auf Netto-Cashflow Automobile und operative Umsatzrendite	A

> **Kernaussage.** Bemerkt wird das Problem durchaus: vom Betriebsrat, von der Presse, von Beratern. Es taucht nur in keiner Messgröße auf, nach der der Vorstand gesteuert oder bezahlt wird. Bekannt ist das Muster seit mindestens 2006.

Mögliche Wurzel (Prüfvorschlag, nicht festgestellt). Grundbedingung ist eine Messlücke: Gemessen wird der Aufwand, nicht die Leistung je Euro. Eingeplante Überschreitungen verlassen das Risikoregister. Verstärkt wird das durch Anreize (Bonus, Aktivierung) und durch Macht und Kultur (Markenautonomie, gefilterte Berichte).

Mechanismus an der Anforderung. Prosa-Anforderungen haben keinen Platz für Abnehmer, Flussgröße und Band. Diese Struktur steht im Kopf erfahrener Entwickler und wird in jeder Abstimmung neu ausgehandelt. Fremdvergabe verursacht die Fehler nicht. Sie verlegt die Köpfe hinter die Vertragsgrenze, und aus dem Gespräch auf dem Flur wird ein Change Request.

Drei Hebel mit Stellautorität:

1. Anforderungsqualität als Abnahmekriterium vor jeder Vergabe messen (Vorstand Entwicklung und Einkauf).
2. Leistungskennzahl F&E je Programm einführen, Plan-Abweichung bleibt im Risikobericht sichtbar (Prüfungsausschuss des Aufsichtsrats, Vorstand Finanzen).
3. Freigabe- und Abstimmungsaufwand mit drei Zahlen messen (Vorstand Entwicklung).

Jeder Hebel ist zugleich ein Test: Er sagt vorher, woran man sieht, ob die vermutete Ursache stimmt, und wann man abbricht.

1 Lesehilfe

Evidenzstufen

Stufe	Bedeutung
A	amtliche Statistik, Register, Pflichtveröffentlichung (Geschäftsbericht, Vergütungsbericht)
B	unabhängige Forschung mit offener Methodik
C	Verbands-Messreihe mit offener Methodik
D	Umfrage, Selbstauskunft, eigene Beobachtung
E	Aussage von Anbieter, Berater oder Verband im eigenen Interesse, Pressemitteilung

Kernaussagen stehen auf A bis C. D und E erscheinen nur als markiertes Indiz, mit dem Interesse des Absenders.

Umrechnung

Reale Werte in Preisen von 2020, deflationiert mit dem Verbraucherpreisindex (VPI, 2020 = 100): 2015 94,5 · 2016 95,0 · 2017 96,4 · 2018 98,1 · 2019 99,5 · 2020 100 · 2021 103,1 · 2022 110,2 · 2023 116,7 · 2024 119,3 · 2025 121,9 (2025 fortgeschrieben mit +2,2 %).

Fremdwährungen zum EZB-Jahresmittel 2024: 163,85 Yen je Euro, 7,7875 Yuan (CNY) je Euro.

Der VPI ist ein Verbraucherindex. Für Ingenieursleistungen wäre ein Lohnindex genauer. Er läge höher, die realen Zuwächse fielen damit etwas kleiner aus. Die Richtung ändert sich nicht.

2 Aufwand und Umfang

Die erste Frage war, ob der Aufwand nur deshalb steigt, weil mehr entwickelt wird. Dafür wird er auf zwei Bezugsgrößen normiert: je verkauftes Fahrzeug und je Modell.

2.1 Volkswagen-Konzern je Fahrzeug

Jahr	F&E Mio. €	Absatz Tsd.	F&E je Fzg. nominal	F&E je Fzg. real	Verwaltung je Fzg. real	Verwaltung in % Umsatz
2016	13.672	10.391	1.316 €	1.385 €	743 €	3,38
2017	13.135	10.777	1.219 €	1.265 €	782 €	3,54
2018	13.640	10.900	1.251 €	1.275 €	825 €	3,74
2019	14.306	10.956	1.306 €	1.313 €	896 €	3,87
2020	13.885	9.157	1.516 €	1.516 €	1.026 €	4,22
2021	15.583	8.576	1.817 €	1.762 €	1.178 €	4,16
2022	18.908	8.481	2.229 €	2.023 €	1.247 €	4,18
2023	21.779	9.362	2.326 €	1.993 €	1.165 €	3,95
2024	≈ 21.000	9.037	2.324 €	1.948 €	1.183 €	3,93
2025	≈ 19.400	9.022	2.150 €	1.764 €	1.140 €	3,90

Quelle: Geschäftsberichte Volkswagen AG (A). F&E 2024 und 2025 gerundet berichtet. Verwaltungskosten nominal: 7.336 Mio. € (2016) → 12.541 Mio. € (2025), +71 %.

Lesart. F&E je Fahrzeug liegt 2025 real 27 % über 2016, mit einer Spitze 2022/23 (+44 %). Die Verwaltung je Fahrzeug liegt real 53 % höher. Der Absatz ist um 13 % gesunken.

Vorbehalt. F&E ist weitgehend Fixkosten. Sinkt der Absatz, steigt jede Kennzahl je Fahrzeug von selbst. Deshalb die zweite Normierung.

2.2 Volkswagen-Konzern je Modell

2016 nennt der Geschäftsbericht 336 Modelle. Für 2026 nennen Konzern und Fachpresse rund 150, Ziel sind 75. Real gerechnet steigt F&E je Modell von 42,8 Mio. € (2016) auf 106,1 Mio. € (2025), Faktor 2,5.

Der Umfang ist also geschrumpft, der Aufwand gewachsen. Das spricht gegen die Erklärung „Wir entwickeln einfach mehr“.

Vorbehalt (offen): Ob „Modell“ 2016 und 2026 gleich abgegrenzt ist, ist nicht belegt. 2016 ist auch von 340 Modellvarianten die Rede. Die Zahl 150 stammt aus Fachpresse und Konzernaussage (E).

2.3 BMW und internationaler Vergleich

BMW: F&E je ausgeliefertes Fahrzeug real 2.434 € (2015) → 2.770 € (2025), +14 %; Spitze 2024 mit 3.105 € (A). Der Anstieg ist deutlich kleiner als bei VW.

Hersteller 2024	F&E	Fahrzeuge	F&E je Fahrzeug
BMW	9.078 Mio. €	2,45 Mio.	3.704 €
Volkswagen	≈ 21.000 Mio. €	9,04 Mio.	2.324 €
BYD (Gesamtkonzern)	54,2 Mrd. CNY = 6.960 Mio. €	4,27 Mio.	1.629 €
Toyota (Geschäftsjahr bis 03/2025)	1,32 Bio. Yen = 8.056 Mio. €	9,36 Mio.	861 €

Nominal, Umrechnung zum EZB-Jahresmittel 2024. Nicht gleich abgegrenzt: Rechnungslegung, Konzernumfang (BYD enthält Elektronik und Batterie), Behandlung aktivierter Entwicklungskosten. Der Vergleich ist ein Hinweis, kein Beweis.

3 Fremdvergabe und Personal

Für die Branche liefert der Stifterverband die einzige Reihe, die interne und externe F&E trennt (Fahrzeugbau, WZ 29, Stufe C).

Jahr	intern Mio. €	extern Mio. €	Anteil extern	Personal VZÄ*	intern je VZÄ real
2016	21.889	9.808	30,9 %	113.865	202,4 Tsd. €
2017	25.656	11.985	31,8 %	126.413	210,5 Tsd. €
2018	27.076	12.638	31,8 %	131.597	209,7 Tsd. €
2019	28.253	13.624	32,5 %	139.331	203,8 Tsd. €
2020	24.417	13.992	36,4 %	134.225	181,9 Tsd. €
2021	26.011	15.054	36,7 %	137.432	183,6 Tsd. €
2022	28.746	15.787	35,5 %	153.300	170,2 Tsd. €
2023	30.346	18.276	37,6 %	157.741	164,8 Tsd. €

Quelle: Stifterverband, Zahlenwerk 2025. VZÄ = Vollzeitäquivalente.

Personal: Bruch in der Reihe. Nominal ergibt die Reihe +39 % Personal von 2016 bis 2023. Das ist zu viel, um plausibel zu sein. Das Zahlenwerk 2019 nennt für 2017 noch 133.830 Personen statt 126.413: Die Reihe wurde revidiert, und der Sprung 2016→2017 (+11 %) ist vermutlich ein Bruch. Belastbar ist 2017–2023 auf gleicher Basis: **+25 %**.

Wer da ist. 2023: 100.908 Forscher, 32.892 Techniker, 23.941 Sonstige. Ob der Zuwachs eher in der Entwicklung oder in Steuerung und Verwaltung (auch IT) liegt, trennt die Statistik nicht. Offen.

Lesart. Die externe F&E wächst mehr als doppelt so schnell wie die interne. Je internem Kopf steht real 19 % weniger Budget zur Verfügung. Der Sprung des Fremdanteils lag 2019→2020.

> Mehr Fremdvergabe heißt mehr Steuerung: Lastenheft, Angebot, Abnahme, Change Request, Eskalation. Jeder unklare Satz im Lastenheft wird an der Vertragsgrenze zu Geld.

4 Verwaltung, Abstimmung, Unmut

4.1 Was Abstimmungszeit kosten kann

Der Umsatzanteil der Abstimmung ergibt sich aus Zeitanteil × Personalaufwandsquote. Bei VW lag die Personalaufwandsquote 2025 bei 15,2 % (49.080 Mio. € Personalaufwand bei 321.913 Mio. € Umsatz, A).

Annahme Zeitanteil Abstimmung	Anteil am Umsatz	VW 2025
10 %	1,5 %	≈ 4,9 Mrd. €
20 %	3,0 %	≈ 9,8 Mrd. €
40 %	6,1 %	≈ 19,6 Mrd. €

Schon 20 % Abstimmungszeit kosten mehr als das gesamte operative Ergebnis von 2025 (8,9 Mrd. €). Eine Angabe von „6 % vom Umsatz“ würde 40 % der gesamten Arbeitszeit voraussetzen.

Rechenbeispiel, keine Messung. Der Zeitanteil ist Annahme. Eine Umfrage nennt 8,7 Stunden je Woche in Meetings (Stepstone, D), das wären gut 20 %. Nicht jedes Meeting ist Abstimmung.

4.2 Abstimmung als Fehlersignal

Sosa, Eppinger und Rowles (2004, B) zeigen: Abstimmung zwischen Teams ist dort legitim, wo eine Schnittstelle im Produkt noch nicht festgelegt ist. Wo Produkt- und Organisationsschnittstellen auseinanderfallen, entsteht Abstimmungsbedarf, der aus dem Produkt nicht nötig wäre.

Damit ist Abstimmung ein Messsignal: Sie zeigt eine Schnittstelle, deren Daten fehlen, oder eine Entscheidung, deren Stellautorität fehlt. Terwiesch und Loch (1999, B, Automobil) fanden je Tag Bearbeitung einer Änderung zwei Wochen Wartezeit.

4.3 Stimmen aus der Belegschaft

- VW-Betriebsratsumfrage, August 2026: Mitarbeiter beschreiben die Kommunikation als „katastrophal“ (D; Absender Betriebsrat, Interesse an Mitbestimmung).
- Stepstone-Umfrage: 8,7 Stunden je Woche in Meetings (D; Absender Jobplattform).

Umfragen belegen Stimmung, nicht Kosten.

5 Anforderungen: Prosa statt Struktur

5.1 Einfluss der Anforderungen

- **NaPiRE** (Méndez Fernández u. a. 2017, internationale Befragungsreihe, B/D): Das am häufigsten genannte Problem sind unvollständige Anforderungen, gefolgt von Kommunikationsmängeln zwischen Team und Kunde, unterspezifizierten Anforderungen und beweglichen Zielen. Die Studie gewichtet auch, welche Probleme zum Scheitern von Projekten führten.
- **Mercedes-Benz** (Ott 2012, B): rund 5.800 Review-Einträge zu Spezifikationen, Lastenhefte mit bis zu 3.000 Seiten.
- **Bosch** (Langenfeld, Post, Podelski 2016, B): 588 Anforderungsdefekte über 4,5 Jahre eines Projekts. 61 % stammen aus Unkorrektheit oder Unvollständigkeit; am teuersten zu beheben sind Unvollständigkeit und Inkonsistenz.
- **MAN Truck & Bus** (Vogelsang, Femmer, Winkler 2016, B): 46 % der untersuchten Review-Befunde betreffen Betriebsmodi, also genau die Bedingung, unter der eine Anforderung gilt.
- **Requirements Smells** (Femmer u. a., B): Formulierungen wie „sufficient accuracy“ oder „as far as possible“ sind maschinell auffindbar und tauchen in Industriespezifikationen regelmäßig auf.

5.2 Der QuRE-Datensatz

QuRE (Femmer, Houdek, Unterbusch, Vogelsang 2025, B) enthält 2.111 Anforderungen aus Mercedes-Spezifikationen für Steuergeräte, Sensoren und Aktoren, in Englisch, frei formuliert. Lizenz CC BY 4.0.

- Die CSV hat genau vier Spalten: id, Anforderungstext, Schwachwort, Bewertung (defect / ok). 633 Einträge sind als fehlerhaft bewertet.
- **Keine Merkmale, kein Systemelement, kein Typ, keine Messwerte.** Ob die Sätze einem Muster folgen, hat niemand ausgewertet.

- Bewertet haben bis zu drei Tester bei Mercedes; die Übereinstimmung zwischen ihnen ist nicht angegeben.
- Die Stichprobe enthält nur Sätze mit Schwachwörtern. Häufigkeiten gelten nicht für Mercedes-Spezifikationen insgesamt.

Beispiel aus dem Datensatz: „The contractor shall be prepared that the dimensions will change in detail.“ Subjekt ist der Auftragnehmer, kein Systemelement. Es gibt kein Merkmal, kein Band, keinen Abnehmer, keinen Auslöser. Der Satz ist eine Absichtserklärung zur Vertragsbeziehung.

Einwand, der vorweg zu entkräften ist: Dass QuRE keine Merkmalsspalten hat, beweist nicht, dass sie im Original fehlen. Im Anforderungswerkzeug hängen meist Attribute am Objekt. Tragend ist der Satz selbst: Ein Attribut am Objekt ersetzt keine Struktur im Satz.

5.3 Satzmuster gibt es seit Jahren

- **EARS** (Mavin u. a. 2009, Rolls-Royce): Bedingung, Auslöser, System, Reaktion. Die Fallstudie zeigt Verbesserungen gegenüber Freitext (B, aber von den Entwicklern der Methode).
- **ISO/IEC/IEEE 29148** und **INCOSE Guide to Writing Requirements**, Version 4 von 2023: Eigenschaften guter Anforderungen und 42 Schreibregeln. Leitfäden, kein Wirkungsnachweis.
- **Vergleich der Schablonen** (Requirements Engineering Journal, 2024, B): Schablonen verbessern verschiedene Qualitätsfaktoren gegenüber Freitext; MASTeR (Rupp) und Boilerplates erreichen die beste Vollständigkeit.

Trotzdem beschreibt das QuRE-Paper die Mercedes-Anforderungen als freie, unbeschränkte Sprache. Die Muster sind bekannt, angewendet werden sie nicht.

5.4 Warum Satzmuster allein nicht reichen

EARS ordnet den Satz. Es verlangt aber nicht, dass Abnehmer, Flussgröße und Band als eigene, verknüpfte Objekte existieren. Der Satz bleibt Text in einem Dokument, an nichts angeheftet. Die Satzschablone des Lebensfähigkeitsmodells verlangt mehr:

> Solange ⟨Lebensphase × Zustand⟩, wenn ⟨Auslöser⟩, liefert ⟨Systemelement⟩ ⟨Flussgröße⟩ ⟨Band am Merkmal⟩ an ⟨Abnehmer⟩.

Jeder Platzhalter ist ein Objekt im Modell, und die Anforderung ist an das Element geheftet, das sie erfüllen muss (Grundsatz 3 der Methode: Zerlegt wird das Produkt, Anforderungen werden angeheftet). Fehlt ein Platzhalter, ist das als Lücke sichtbar, bevor das Lastenheft das Haus verlässt.

Folgerung: EARS ist notwendig, aber nicht hinreichend. Das freie Schweben der Anforderung beendet erst das Anheften an die Struktur.

5.5 Austausch über die Vertragsgrenze

- Liebel u. a. (2018, B): Beim Austausch zwischen Hersteller und Zulieferer fehlen klare Verantwortungen, und es entstehen Lücken zwischen den Abstraktionsebenen.
- ReqIF ist der Austauschstandard. Ein Benchmark von prostep ivip (2025, C/E, Verband der Anwender und Anbieter) meldet weniger Austauschfehler als zuvor. ReqIF transportiert Text und Attribute, nicht die Bedeutung des Satzes.

5.6 Mechanismus (Hypothese)

Die fehlende Struktur wird durch Menschen ersetzt. Abnehmer, Flussgröße und Band stehen im Kopf erfahrener Entwickler und werden in Abstimmungen jedes Mal neu ausgehandelt. Das funktioniert, solange diese Köpfe im eigenen Haus sitzen.

Fremdvergabe verlegt die Köpfe hinter die Vertragsgrenze. Aus dem Flurgespräch wird ein Change Request mit Angebot und Eskalation. Die Lücke war immer da, mit der Fremdvergabe wird sie teuer. Das passt zu extern +86 % und Verwaltung je Fahrzeug +53 %, ohne dass Fremdvergabe selbst als Fehlerursache behauptet wird.

In der Sprache der Methode ist das ein Substanzverzehr: Gemessen wird die Entnahme (vergebenes Volumen, geschlossene Aufträge), nicht der Bestand, aus dem sie kommt (internes Systemwissen). Die interne F&E je Kopf sinkt real um 19 %.

6 Software im Feld

- ID.3: Unfertige Software verzögerte die Auslieferung, ausgeliefert wurde ohne vollen digitalen Umfang. Erstes Update über die Luftschnittstelle (OTA) im Juli 2021.
- 2022 mussten rund 200.000 ID.3 und ID.4 in die Werkstatt (Software 2.4, neue 12-V-Batterie), damit das OTA-Update auf 3.0 überhaupt möglich wurde. „Over the air“ wurde zu „über das Autohaus“.
- Noch 2023 und 2025 gibt es Werkstattanweisungen bei der US-Behörde NHTSA zu fehlgeschlagenen OTA-Updates (A).
- Rückrufe in Deutschland: KBA 2,84 Mio. Fahrzeuge 2024 (+48 %), 1,78 Mio. 2025 (−37 %) (A). ADAC: 532 Rückrufaktionen 2025 gegenüber 185 im Jahr 2010 (C).
- Anteil der Software an US-Rückrufen 12,3 % → 23 % (Kanzlei für Produkthaftung, E).

Kausalität: der Zeitversatz-Test

Fremdvergabe und Rückrufe fallen zeitlich zusammen. Das ist kein Beleg: Erst kommt die Vergabe, dann die Serie, dann das Feld. Wenn Fremdvergabe zu Rückrufen beiträgt, muss der Anstieg mit Versatz kommen. Der Sprung des Fremdanteils lag 2019→2020; betroffen müssten vor allem Fahrzeuge der Baujahre ab etwa 2023 sein.

Prüfbar mit der KBA-Rückrufdatenbank, die je Aktion „Baujahr von/bis“ führt: Fremdanteil im Jahr t gegen Rückrufe der Baujahre t+3 bis t+5. Offen, lokal durchführbar.

7 Warum der Vorstand nicht nervös wird

7.1 Es ist kein Wissensproblem

Grimm (2003, B) zitiert Experten von DaimlerChrysler: 80 % der künftigen Innovationen über Elektronik, davon 90 % über Software. Broy u. a. (2006/07, B): mehr als 80 % der Innovationen kommen aus Rechnersystemen. McKinsey (E): Die Software-Komplexität wuchs in zehn Jahren um den Faktor 4, die Produktivität nur um 1,0 bis 1,5.

Die Frage ist also nicht, warum es niemand weiß, sondern warum niemand danach steuert.

7.2 Der Risikobericht ist so gebaut, dass er es nicht zeigt

- VW stuft das Entwicklungsrisiko in den Berichten 2020 bis 2024 identisch ein: Eintrittswahrscheinlichkeit hoch, Ausmaß mittel (A). Berichtet wird ab einem Score von 40 oder einem möglichen Schaden ab 1 Mrd. €. Mercedes führt das F&E-Risiko 2024 und 2025 als „Medium“.
- Im selben Zeitraum schätzte McKinsey (2022, E) die Mehrkosten der Cariad-Architektur auf rund 3,5 Mrd. € bis 2026 und 9,2 Mrd. € bis 2039, dazu 2,5 bis 3 Mrd. € durch die Verzögerung bei Porsche. Der Vorstand kannte die Größenordnung.
- **Mechanismus:** Alle drei Hersteller nehmen eingeplante Risiken aus dem Register. Eine Überschreitung, die ins Budget übernommen wird, ist kein Risiko mehr, sondern Plan. Das Register ist eine Restmenge.
- Die einzige veröffentlichte F&E-Kennzahl ist die F&E-Quote: Aufwand im Verhältnis zum Umsatz. Sie misst, was hineingeht. Eine Kennzahl für Leistung je F&E-Euro steht in keinem der Berichte.

7.3 Was die Vergütung misst

- Der VW-Vorstandsbonus bemisst sich je zur Hälfte am Netto-Cashflow Automobile und an der operativen Umsatzrendite (A). Wie wirksam F&E-Geld eingesetzt wird, geht nicht ein.

- VW aktiviert 48,1 % der Entwicklungskosten (2025, A). Das verschiebt Aufwand in spätere Jahre.
- Für 2025 gab es den Vorwurf, Entwicklungskosten seien ins Folgejahr verschoben worden; der Cashflow lag laut Spiegel knapp über der Schwelle für den vollen Bonus. Der Konzern bestreitet das (D/E, nicht belegt). Der Mechanismus selbst ist offen sichtbar.

7.4 Macht und Organisation

- Cariad: Abweichungen im Entwicklungsstand, unterschiedliche Vorstellungen der Marken, Probleme bei der Kompatibilität.
- Der Betriebsratsvorsitzende Cavallo fordert, den „Markenegoismus“ einzudämmen (E, Arbeitnehmerseite). Das ist das Thema Stellautorität: Wer darf eine Schnittstelle festlegen?

7.5 Kultur und Wahrnehmung

- INSEAD-Fallstudie: lang bestehende Kultur der Angst, verstärkt durch dominierende Gruppen im Aufsichtsrat. Darden-Fallstudie: Wer Winterkorn berichtete, sorgte vorher für gute Nachrichten (B). Beides betrifft die Zeit vor 2015 und darf nicht als heutiger Zustand gelten.
- Eine Studie aus Edinburgh beschreibt nach dem Dieselskandal einen Wandel hin zum offenen Melden von Fehlern (B).
- Vaughan (1996, Challenger, B) nennt den Mechanismus „normalization of deviance“: Abweichungen werden Jahr für Jahr umgedeutet, bis sie normal wirken. Konzept, kein Beleg für VW heute.

> Die Aufmerksamkeitsfalle (Pflicht 5 des Lebensfähigkeitsmodells): Auf laute Ereignisse wird schnell reagiert, auf schleichende Signale jahrelang nicht. Ein jährlicher Anstieg von einigen Prozent, jedes Jahr begründet, ist das leiseste Signal, das es gibt.

8 Prüfvorschlag zur Wurzel

Die Ursache lässt sich von außen nicht feststellen; die Daten liegen beim Hersteller. Deshalb ein Prüfvorschlag statt einer festgestellten Ursache.

Belegte Abweichung. Bei VW 2016–2025: F&E je Fahrzeug real +27 %, je Modell Faktor 2,5, Verwaltung je Fahrzeug real +53 %, Absatz −13 %. In der Branche: externe F&E +86 %. Die Risikoeinstufung bleibt über fünf Jahre unverändert.

Mögliche Ursachen (geschätzt):

	Ursache	Woran man sie erkennt (Trennschärfe)
H0 echte Komplexität	Elektrifizierung, Software und Verbrenner parallel: Mehraufwand ist legitim	Aufwand je Funktion oder je Steuergerät bleibt flach; Wettbewerber mit ähnlicher Technik liegen ähnlich. Toyota und BYD liegen je Fahrzeug deutlich niedriger, das spricht gegen H0 allein.
H1 Messlücke	keine Kennzahl für Leistung je F&E-Euro; eingeplante Überschreitungen verlassen das Register	tritt bei allen Herstellern gleich auf, unabhängig von der Kultur; Einstufung bleibt trotz bekannter Milliarden-Überschreitung konstant
H2 Anreiz	Bonus auf Cashflow und Marge, Aktivierung	Häufung von Verschiebungen kurz vor Jahresende und nahe an Bonusschwellen
H3 Macht und Kultur	Markenautonomie, Gremien, gefilterte Berichte	Unterschiede zwischen Herstellern mit unterschiedlicher Führungsstruktur bei gleicher Technik; hohe Abstimmungsquote auch bei gut spezifizierten Umfängen

Prüfschritt. Den F&E-Zuwachs je Programm zerlegen: Plan bei Freigabe (eingefroren) gegen Ist; Zahl der Change Requests je 100 Anforderungen; Anteil Fremdvergabe; Anforderungsqualität der Lastenhefte als Stichprobe von 200 Sätzen je Lastenheft, ausgezählt nach den Bausteinen der Satzschablone, mit zwei Prüfern (Übereinstimmung mindestens 0,7).

Wer die Daten hat. Der Prüfungsausschuss des Aufsichtsrats, als Auftrag an die interne Revision. Daten liefern Entwicklung und Einkauf.

Frist. Sechs Monate.

Abbruch. Die Daten lassen sich nicht zusammenstellen. Dann ist das selbst ein Befund: Der Konzern kann seinen größten Zukunftsposten nicht erklären.

Einschätzung nach der Recherche. H1 ist die Grundbedingung, H2 und H3 verstärken. H0 erklärt einen Teil, aber nicht den Faktor 2,5 je Modell bei schrumpfender Modellzahl.

9 Lösungswege nach dem Lebensfähigkeitsmodell

Das Lebensfähigkeitsmodell prüft jedes zweckgerichtete System mit fünf Pflichten. Hier angewendet auf die Entwicklung eines Herstellers.

Pflicht	Befund heute	Soll
1 Zweck: wofür, für wen?	Gesteuert wird der Aufwand (F&E-Quote). Ein Adressat der Entwicklung ist in keiner Kennzahl benannt.	Zweck mit Adressat: Funktion in Serie beim Kunden, in der Werkstatt, im Werk. Auf der Ebene der Anforderung: der Abnehmer.
2 Idealzustand mit Zahl und Datum	Überschreitungen werden neu eingeplant; der alte Plan verschwindet.	Plan bei Freigabe einfrieren. Die Lücke zwischen eingefrorenem und lebendigem Stand bleibt sichtbar.
3 Messmittelfähigkeit	Gemessen wird die Entnahme (Aufwand, vergebenes Volumen). Leistung je Euro und Anforderungsqualität werden nicht gemessen. Der Bestand (internes Systemwissen) wird nicht geführt.	Vital-Größen: F&E je Modell real gegen Plan; Anteil vollständiger Anforderungssätze (Übereinstimmung ≥ 0,7); internes Systemwissen als Bestand.

Pflicht	Befund heute	Soll
4 Schwachpunkt	Das Register ist eine Restmenge; die Einstufung bleibt konstant.	Je Programm einzeln, schwächstes oben, nie mitteln.
5 Beweglichkeit	Bekannt seit 2006; Software-Probleme im Feld 2020 bis 2025; Freigaben mit langer Wartezeit.	Reaktionszeit messen: von der ersten Abweichung bis zur Entscheidung, gegen die Zeit, die das Marktfenster noch hält.

Wurzel nach der Methode. Ein Kettenglied ist Wurzel, wenn dort eine Pflicht nachweislich verletzt ist und es einen Hebel mit Verantwortlichem, Stellautorität und Wirksamkeitskriterium gibt. H1 erfüllt beides: Pflicht 3 ist verletzt, der Hebel liegt beim Prüfungsausschuss und beim Vorstand Finanzen. An der Anforderung ist Pflicht 1 verletzt (kein Abnehmer), der Hebel liegt beim Vorstand Entwicklung.

9.1 Vier Maßnahmen, jede als Test

Die ersten beiden Maßnahmen sind zugleich der Prüfschritt aus Abschnitt 8. Sie verändern keine Struktur, sie messen. Die vierte ist erst zulässig, wenn der Prüfschritt H2 bestätigt.

M1 Lastenheft-Abnahme vor der Vergabe

- **Zuständig:** Vorstand Entwicklung und Einkauf.
- **Maßnahme:** In zwei bis drei Pilotvergaben werden Anforderungen in der Satzschablone geschrieben und an das Systemelement geheftet. Abnahmekriterium vor Vergabe ist der Anteil vollständiger Sätze; Zielwert vorab festlegen (Vorschlag: 90 %).
- **Test:** Wenn Prosa den Abstimmungsbedarf verursacht, sinken Klärungsrunden und Change Requests je 100 Anforderungen im Pilot gegenüber vergleichbaren Vergaben. Messbar in der Change-Request-Statistik, innerhalb eines Musterstands (12 bis 18 Monate).
- **Abbruch:** Kein Unterschied. Dann liegt die Ursache eher in der Organisation (H3), und M3 geht vor.
- **Preis:** Mehr Aufwand vorn beim Schreiben; Lieferanten müssen das Format mittragen.

M2 Steuergröße statt Aufwandsquote

- **Zuständig:** Prüfungsausschuss des Aufsichtsrats, Vorstand Finanzen.
- **Maßnahme:** Je Programm der eingefrorene Plan bei Freigabe gegen Ist, dazu F&E je Modell real. Eine eingeplante Überschreitung bleibt im Risikobericht als Abweichung stehen und verschwindet nicht in der Planung.
- **Test:** Wenn die Messlücke (H1) die Reaktion verhindert, verkürzt sich die Zeit von der ersten Abweichung über 10 % bis zur Entscheidung (Umplanung, Stopp). Messbar an den Gremienprotokollen, innerhalb von zwei Berichtszyklen.
- **Abbruch:** Die Abweichung ist sichtbar, die Reaktionszeit bleibt. Dann wirken Anreiz oder Macht (H2, H3).
- **Preis:** Unbequeme Zahlen im Bericht; Diskussion mit Kapitalmarkt und Prüfern.

M3 Freigabe und Abstimmung mit drei Zahlen messen

- **Zuständig:** Vorstand Entwicklung.
- **Maßnahme:** Je Änderung erfassen: Zahl der Freigabeinstanzen, Wartezeit je Instanz, Ablehnungsquote je Instanz über mindestens 100 Änderungen. Liegt die Ablehnungsquote bei null, hat die Instanz kein Kriterium. Sie verliert das Stopp-Recht und wird informiert.
- **Test:** Wenn Freigaben ohne Kriterium Totzeit erzeugen, sinkt die Durchlaufzeit von Änderungen, ohne dass die Fehler im Feld steigen. Innerhalb von 12 Monaten.
- **Abbruch:** Die Feldfehler steigen. Dann hatte die Instanz ein Kriterium, das nicht deklariert war, und wird mit diesem Kriterium wieder eingesetzt.
- **Preis:** In der Übergangszeit sichtbar höheres Fehlerrisiko. Das wirkt nur, wenn der Vorstand die Asymmetrie selbst aufhebt: Wer freigibt, haftet für den Verzug wie für den Fehler.

M4 Vergütung (nur nach Bestätigung von H2)

- **Zuständig:** Aufsichtsrat, Vergütungsausschuss.
- **Maßnahme:** Eine Bonuskomponente auf Entwicklungsleistung, zum Beispiel F&E je Modell real gegen Plan oder Software-Ausfälle im Feld je 1.000 Fahrzeuge.
- **Test:** Die Häufung von Verschiebungen am Jahresende verschwindet.

Reihenfolge. M2 und M1 zuerst: billig, ohne Strukturänderung, und sie liefern den Prüfschritt. M3 parallel. M4 erst nach Befund.

Grenze der Methode: Sie macht sichtbar und prüfbar. Entscheiden und den Preis tragen müssen die Zuständigen.

10 Preis des Nichthandelns

- Cariad-Mehrkosten laut McKinsey bis 2039 rund 9,2 Mrd. €, dazu 2,5 bis 3 Mrd. € bei Porsche (E).
- Werkstattaktion für rund 200.000 Fahrzeuge, damit ein Update über die Luft überhaupt möglich wird.
- Jeder Prozentpunkt Abstimmungszeit kostet bei VW rund 0,5 Mrd. € im Jahr (Rechenbeispiel aus 4.1).

Messbar ist der Preis in Größen, die der Hersteller selbst erhebt: Change Requests je 100 Anforderungen, Durchlaufzeit von Änderungen, Software-Ausfälle im Feld, Abweichung Plan gegen Ist je Programm.

Anhang: QuRE nach der Satzschablone auszählen

Näherung mit einfachen Regeln am englischen Text. Die Anteile gelten nur für die Teilmenge mit Schwachwörtern.

Baustein der Schablone	Prüfbares Zeichen im Text
Wenn-Block (Gültigkeit und Auslöser)	Satz beginnt mit „When / While / If / As long as / In case“
Verbindlichkeit	„shall / must / should / will“, getrennt gezählt
Band am Merkmal	Zahl mit Einheit (ms, V, A, °C, km/h, %) und Grenzwort (max., min., ≤, between)
Subjekt = Systemelement	„The ⟨Nomen⟩ shall“, nicht „It / The contractor / The system“
Abnehmer und Flussgröße	„to / at / from ⟨Nomen⟩“ nach dem Verb, nur grob prüfbar

Nicht prüfbar ohne Zusatzdaten: Verbkatalog (Abgleich mit eigener Liste nötig) und Systemelement (keine Bauteilliste im Datensatz).

Quellen

- Volkswagen AG: Geschäftsberichte 2016–2025, Vergütungsberichte, Risikoberichte 2020–2024. (A)
- BMW AG: Geschäftsberichte 2015, 2016, 2024, 2025. (A)
- Mercedes-Benz Group: Risikoberichte 2024, 2025. (A)
- Stifterverband: FuE-Datenreport, Zahlenwerk 2025 und 2019, Wirtschaftszweig 29. (C)
- Statistisches Bundesamt: Verbraucherpreisindex. EZB: Referenzkurse, Jahresmittel 2024. (A)
- Kraftfahrt-Bundesamt: Rückrufstatistik 2024, 2025; Rückrufdatenbank. NHTSA: Technical Service Bulletins 2023, 2025. (A)
- ADAC: Rückrufstatistik 2025. (C)
- Toyota Motor Corporation, Geschäftsjahr bis März 2025; BYD, Geschäftsbericht 2024. (A)
- Méndez Fernández, Wagner, Kalinowski u. a.: Naming the Pain in Requirements Engineering. Empirical Software Engineering, 2017. (B/D)
- Ott: Defects in natural language requirement specifications at Mercedes-Benz. RE 2012. (B)
- Langenfeld, Post, Podelski: Requirements Defects over a Project Lifetime. REFSQ 2016. (B)

- Vogelsang, Femmer, Winkler: Take Care of Your Modes! REFSQ 2016. (B)
- Femmer, Houdek, Unterbusch, Vogelsang: Description and Comparative Analysis of QuRE. REW 2025; Datensatz Zenodo 10.5281/zenodo.15656471, CC BY 4.0. (B)
- Femmer u. a.: Rapid quality assurance with Requirements Smells. Journal of Systems and Software, 2017. (B)
- Mavin, Wilkinson, Harwood, Novak: Easy Approach to Requirements Syntax (EARS). RE 2009. (B, Entwickler der Methode)
- INCOSE: Guide to Writing Requirements, Version 4, Juli 2023. ISO/IEC/IEEE 29148. (Leitfaden)
- Vergleich von Satzschablonen für Anforderungen. Requirements Engineering Journal, 2024, doi 10.1007/s00766-024-00427-0. (B)
- Liebel u. a.: Organisation und Kommunikation zwischen Hersteller und Zulieferer im Requirements Engineering, 2018. (B)
- prostep ivip: ReqIF-Benchmark 2025. (C/E)
- Sosa, Eppinger, Rowles: The Misalignment of Product Architecture and Organizational Structure. Management Science, 2004. (B)
- Terwiesch, Loch: Managing the Process of Engineering Change Orders. Journal of Product Innovation Management, 1999. (B)
- Broy u. a.: Engineering Automotive Software. Proceedings of the IEEE, 2007; Grimm: Software Technology in an Automotive Company, ICSE 2003. (B)
- Vaughan: The Challenger Launch Decision, 1996. (B)
- INSEAD- und Darden-Fallstudien zu Volkswagen; Studie der University of Edinburgh zur Fehlerkultur nach 2015. (B)
- McKinsey: Analyse zur Cariad-Architektur 2022, berichtet in der Presse; Software-Komplexität und Produktivität. (E)
- Spiegel 2026 zum Vorwurf verschobener Entwicklungskosten; Stellungnahme des Konzerns. (D/E)
- Stepstone: Umfrage zu Meetings; VW-Betriebsrat: Umfrage August 2026. (D)
- Arends: Lebensfähigkeitsmodell, Methodenbeschreibung v3, 2026; Delta „Risikoregister als Restmenge“, Oktober 2026. (eigene Methode)