



Manufacturing Data Science

Chia-Yen Lee, Ph.D. (李家岩 博士)

Department of Information Management (資訊管理學系)
National Taiwan University (國立台灣大學)

□ Education

- Ph.D, 工業與系統工程, Texas A&M University, USA
(Major: Operations Research 作業研究/運籌學)
- M.S., 工業工程與工程管理, 國立清華大學
- B.S. & B.B.A., 應用數學暨資訊管理, 國立政治大學

□ Experience

- 教授，國立台灣大學資訊管理學系
- 教授兼所長，國立成功大學資訊工程學系暨製造資訊與系統研究所
- 副編輯，IEEE Transactions on Automation Science and Engineering (SCI)
- 工業工程學會秘書長、工業局新興技術專家顧問、半導體廠科技顧問

□ Award

- 台大管理學院翁肇喜先生講座(2022)、科技部傑出研究獎(2022)
- IEEE Senior Member(2021)、呂鳳章先生紀念獎(2019)
- 美光教師Micron Teacher Award (2018)
- 李國鼎科技與人文講座研究獎 (2018)、科技部吳大猷先生紀念獎 (2017)

□ Research Interest

- 製造數據科學、智慧型製造系統、生產力與效率分析、排放權交易、多目標決策



□ Introduction

- MDS course supports students learning how to apply artificial intelligence (AI), machine learning, data science (DS) techniques to improve the effectiveness and efficiency of the **manufacturing systems**. MDS integrates the knowledge domains of the **information, engineering, and management**. Encourage students to solve the real problem **systematically** using the design of analytics, from descriptive, diagnostic, predictive, prescriptive to automating, for successfully enhancing decision quality.

□ Objectives

- Learn the **statistical learning and optimization** methodologies for intelligent manufacturing systems
- Create a **prototype model** to solve the problem in real setting related to manufacturing or service systems
- Develop the **research** skills and prepare a **analytic project report**

□ Learning Requirement:

- It's better to have prerequisite courses: (1) probability and statistics; (2) operations research
- Python programming skills
- Students need to read literature and develop analytical model for term project

□ Challenges

- How to find the problem and identify the right methodology to solve the problem in manufacturing systems?
- Effectiveness vs. Efficiency
 - From Organization Management Perspective (Drucker, 1977)
 - Do the right thing → Effectiveness
 - Do the thing right → Efficiency
 - From Production System Perspective (Lee and Johnson, 2015)
 - Demand over Output → Effectiveness
 - Output over Input → Efficiency
- Optimization Theory and Statistics
- IT Capability & Programming Skills (eg. Python only)

Drucker, P., 1977. An Introductory View of Management. Harper College Press, New York.
Lee, C.-Yen, and A. L. Johnson, 2015. Effective Production: Measuring of the Sales Effect using Data Envelopment Analysis. Annals of Operations Research, 235 (1), 453–486.

□ Textbook

- 李家岩、洪佑鑫 (2022)，製造數據科學，前程文化。Lecture Notes

□ Reference

- Hastie, T., R. Tibshirani, and J. Friedman (2009), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer.
- Hillier, F. S., Lieberman, G. J. (2010), Introduction to Operations Research, 9th ed., McGraw-Hill, New York.
- Hopp, W. and M. Spearman (2011), *Factory Physics*, 3rd ed., Waveland Press.
- Montgomery, D. C. (2013), Introduction to Statistical Quality Control, 7 ed.: John Wiley & Sons, Inc.
- Nahmias, S. (2008), *Production and Operations Analysis*, 6th ed., McGraw-Hill/Irwin.
- Pinedo, M. L. (2016), Scheduling: Theory, Algorithms, and Systems, 5th edition, Springer-Verlag New York.

國立高雄科技大學講座教授/逢甲大學講座教授/國立中興大學前副校長 周至宏

全球產業界關心之智慧機械/智慧製造/工業4.0的研發議題，涵蓋生產系統、品質保證、製程優化與數位建模，應用與整合實驗設計、Inductive及Deductive人工智慧等技術，提供所需之解決方案；上述技術可說是建立在數據科學方法的基礎上。李家岩教授深研製造數據科學實務，在數據科學之學術理論研究與產業實務應用皆有優異且扎實的成就，以其多年的研究及教學經驗撰寫此書，除了可以提供不同產業的主管參考外，也可以作為培育智慧機械/智慧製造/工業4.0領域人才的重要教材。

國立臺灣大學財務金融系教授兼管理學院院長 胡星陽

數據科學是以統計、資訊科學為基礎方法，應用到各領域的跨領域整合。「不積跬步，無以至千里；不積小流，無以成江海。」數據科學所有的應用都必須從基礎方法的掌握開始。李家岩、洪佑鑫兩位作者出版的這本新書《製造數據科學》提供給讀者的就是基礎學理與進階方法。以製造為經、數據科學為緯，讓讀者從奠定期基礎開始，進而得以客製化應用，提升本職學能。這本書內容充實，做為案頭的參考書也很適合，讀者也一定能從本書得到豐富的收穫。

國立成功大學資訊工程學系教授/中華民國人工智慧學會理事長 高宏宇

近年在AI浪潮下，在金融、生醫、交通與製造場域中各式創新、複雜與威力強大的人工智能解決方案已有十足的發展。「內容為王」到「數據為王」都說明了擁有資料就掌握了成功的關鍵。本書詳細介紹資料科學常用的工具與方法，包含清洗、特徵工程、資料降維與資料擴增，到訓練模型的超參數最佳化與近來很重要的可解釋性議題。並在製造場域的資料取得實證，並輔以不同的學習策略。此書涵蓋資料科學工具本質與善用工具產生價值藝術，實可讓讀者成為資料科學藝術家。

華邦電子(股)公司總經理 陳沛銘

數據在新一代的製造中扮演非常重要的角色，在先進工廠例如半導體晶圓製造Fab中，每秒產生數以百萬筆計的資料與圖像，作為故障預測、良率改善、生產力提升的依據，從而邁進智慧工廠。哪些資料必須被蒐集、記錄、整理及運用，是晶圓廠的專業及經驗判斷(Domain Knowledge)；而如何有效地整理和分析數據，提供即時生產修正或改善依據，則是數據科學(Data Science)的專業領域。李家岩教授與洪佑鑫先生合著的《製造數據科學》探討數據科學的基礎知識與先進數據處理方法，作為學習教科書或日常應用的參考，皆可得心應手。

台灣人工智慧學校校務長 蔡明順

數據科學是建構人工智慧的基礎，人工智能是推動智慧製造的關鍵，智慧製造正是第四波工業革命的實現。這本《製造數據科學》從理論基礎、技術應用、架構舉證，最後到製造業場景，完整清晰地貫通數據分析的精髓，精彩的問題討論及案例，都是製造產業現場的具體實證。作者李家岩教授學養及實務經驗豐富，是新一代人工智能領域的巨星，本書注定要成為數據科學教科書的聖經，智慧製造界的扛鼎之作。

台達電子機電事業群總經理 蔡清雄博士

從產線製程到用戶歷程的品質成本管控一直是企業經營重點。當從監控設施取出的數據與產品品質的正相關連結度不高時，即使投入更多監控設施，品質風險未必有效降低。這本《製造數據科學》提供實務面和理論面的對應解決方法，將監控量測成本和品質風險成本得以更有效取得兩者最佳化的利益韌韻。同時透過製造數據來提早發覺可能的製程設備或半成品的不良問題，做超前的防範工作，避免不良產品的出廠。李家岩教授和洪佑鑫先生的合著，從「製造系統分析與管理」到「故障預測與健康管理」，清楚的章節主題，都有深入淺出的精闢闡述，讀者可以很容易找到在各條件下的解決方法，是值得收藏且一生受用。

國立成功大學講座教授/智慧製造研究中心主任/中華民國自動化科技學會前理事長 鄭芳田

作者李家岩老師多年開設課程「智慧型製造系統(Intelligent Manufacturing Systems)」，回響熱烈。如今將教學與產學經驗，彙集成書，為教育推廣盡一份心力。本書介紹製造系統與數據科學的原理及進階方法，在工業自動化與大數據的時代，整合資訊科技(IT)與工業工程(IE)的思維，相信為邁向工業4.1零缺陷的智慧製造系統，奠定期厚實的學理基礎。

國立清華大學講座教授/科技部人工智慧製造系統研究中心主任/

中國工業工程學會理事長 簡禎富

博士指導教授Francois Saintfort在我畢業時說「當教授最欣慰的是能指導可以超越自己的學生」，非常高興和十二萬分榮幸推薦李家岩教授與洪佑鑫先生合著《製造數據科學》，本書基於作者傑出的研究貢獻、精彩的教學模式和豐碩的產學合作成果，深入探討數據科學理論、製造實務洞察與人工智慧、大數據分析等技術的結合，內容豐富詳實，可以組合不同章節作為相關課程的教科書，也是產業界人士自修和必備的工具書。

(依姓氏筆畫排序)



製造數據科學

邁向智慧製造與數位決策

李家岩、洪佑鑫著



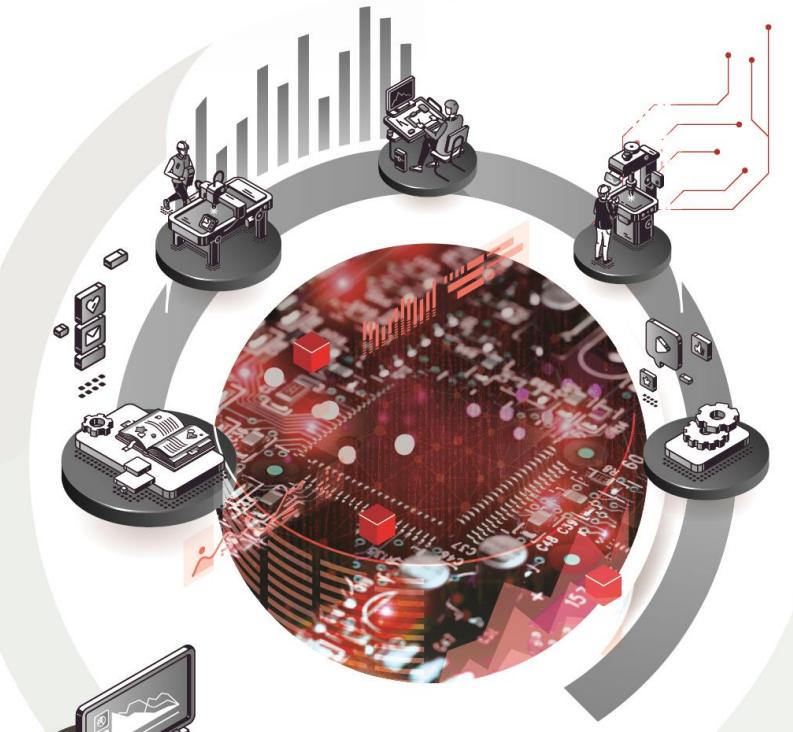
製造數據科學

邁向智慧製造與數位決策

李家岩、洪佑鑫(2022)

Chia-Yen Lee, Ph.D. 7

▶ 李家岩、洪佑鑫 著



- 第一章 製造數據科學
- 第二章 製造系統分析與管理
- 第三章 數據科學基礎與模型評估
- 第四章 數據科學分析架構與系統運算決策
- 第五章 數據預處理與製造數據特性
- 第六章 線性分類器
- 第七章 無母數迴歸與分類
- 第八章 決策樹與集成學習
- 第九章 特徵挑選與維度縮減
- 第十章 類神經網路與深度學習
- 第十一章 集群分析
- 第十二章 特徵工程、數據增強與數據平衡
- 第十三章 故障預測與健康管理
- 第十四章 可解釋人工智慧
- 第十五章 概念漂移
- 第十六章 元啟發式演算法
- 第十七章 強化學習

藍：老師課堂講授
綠：學生自學

- 附錄A 線性迴歸
 - 附錄B 支持向量機
 - 附錄C 統計製程管制與先進製程控制
 - 附錄D 超參數最佳化
-
- 應用涵蓋
產能規劃、瑕疵檢測、製程監控與診斷、機台保養、需求預測、生產排程、電腦視覺、自動光學檢測、原料價格預測與採購等

□ Course Grading Scheme

- 4 Assignments 50%
- Workshop attendance 10% (11/26 智慧製造與產業實務研習營)
- Research Project 40% (Analytics with System Demo)
 - or Comprehensive Exam

□ Project Grading Scheme (40%)

- PIP保密數據不可使用，後果與風險自行承擔。
- Proposal 10%
- Presentation 40%
 - 後一組問前一組問題
 - 同儕互評(其他組別評分後取中位數)
- Final Report 50%

□ Course Website

- NTU Cool (<https://cool.ntu.edu.tw/login/portal>)
- Github: <https://github.com/PO-LAB/Manufacturing-Data-Science>

2020 Intelligent Manufacturing and Industry Practice Workshop 智慧製造與產業實務研習會

研習會議程

09:00~09:20 註冊與報到

09:20~09:30 開幕致詞 – 國立成功大學 蘇慧貞 校長
電機資訊學院 許渭州 院長

09:30~10:20 講者：漢翔航空工業 吳天勝 研發長
人工智慧與作業研究在航空製造業的應用

10:20~10:40 大合照、茶敘與交流 (Break)

10:40~11:30 講者：旺宏電子 陳瑞坤 副總經理
智慧製造在半導體業的應用

11:30~12:20 講者：新漢科技 林弘洲 總經理
智造賦能、工業4.0關鍵技術與對標應用

12:20~13:30 中餐時間

13:30~14:20 講者：成功大學 陳曉民 特聘教授
智能化企業再造

14:20~15:10 講者：友達光電 王德弘 處長
AI驅動新智造

15:10~15:30 茶敘與交流 (Break)

15:30~16:20 講者：富強鑫精密工業 林宗彥 部經理
挖掘射出成型工藝數據的力量

16:20~16:50 論壇與討論
引言人：國立成功大學 李家岩 教授
與談人：講者群

16:50 閉幕、交流討論與賦歸

DATE
1/18 (六) 09:00~16:50

聯絡電話：06-2757575 擴 34205 余小姐
聯絡信箱：em63900@email.ntku.edu.tw

報名費用：NTD 1000
(含講義、午餐、區塊鏈研習證書)
報名連結：<https://reurl.cc/Qpxq19>

Intelligent Manufacturing and Industry Practice Workshop 智慧製造與產業實務研習會

2022 / 01 / 08 (六)

0900 – 0920 註冊與報到

0920 – 0930 開幕致詞
國立臺灣大學管理學院院長 胡星陽
科技部工業工程與管理學門召集人 范書愷

0930 – 1010 淺談數位轉型看台灣智慧製造
經濟部工業局電子資訊組 呂正欽 副組長

1010 – 1050 應用AI與數位雙生技術
於智慧製造之產業案例分享
科技部自動化學門召集人 蔡孟勳 特聘教授

1050 – 1105 茶敘與交流

1105 – 1145 Smart Manufacturing in Semiconductor Wafer Fab
華邦電子 李馥源 副總監

1145 – 1205 智慧製造人才多元培訓模式說明
(推廣/產學媒合/顧問導入/產業鏈共創)
經濟部工業局智慧電子學院 劉依瑜 規劃師

1205 – 1220 Q&A、產學交流、賦歸

連絡人 江小姐
聯絡信箱 r10725034@ntu.edu.tw
報名連結

主辦單位 國立臺灣大學管理學院、國立臺灣大學資訊管理學系
國立臺灣大學資訊管理學系
Dept. of Information Management
National Taiwan University

協辦單位 國立成功大學製造資訊與系統研究所 國立成功大學資訊工程學系
MOST
Ministry of Science and Technology

協辦單位 國立成功大學國際產學聯盟
AIMS
AIMS Artificial Intelligence Manufacturing System Research Center

中國工業工程學會 **社團法人中國工業工程學會**

□ 朱彬祺

- Expertise
 - Data Science
 - Prognostic and Health Management (PHM)
- Email: r10725002@ntu.edu.tw

□ 莊芯瑜

- Expertise
 - Operations Research and Vehicle Routing Problem
 - Preventive Maintenance (PM) Scheduling
- Email: r10725014@ntu.edu.tw

□ Office Hour: by appointment

- 講義內容會以重點說明方式進行
 - 有些內容會請同學參考或自行閱讀即可(可與教科書相互對照)
- 課程需分組
 - 四人一組，盡量不同科系
 - 專案需要跟製造業或服務業有關
- 課程作業基本上沒有標準答案
 - 有答案寫法，但是過沒多久就不新了...甚至效果不好了..(IT進步很快)
- 加簽就到今天第一堂課中午前
- 遠距上課若斷線請不用擔心
- 登入時請用自己的姓名與學號

□ Prerequisite

- Probability and Statistics
- Operations Research
- Python

□ 是研究所課程

- 有一定的loading...

□ 是跨領域

- 有一定的學習挑戰
- 與小繁雜...

□ 製造業並不是那麼美好與理想...(但其實各行各業都一樣~)

□ 整學期上完大概僅佔XX%吧..但掌握了關鍵**基礎學理與態度**..





Manufacturing Data Science

Manufacturing Data Science (第 1 章 製造數據科學)

Chia-Yen Lee, Ph.D. (李家岩 博士)

Department of Information Management (資訊管理學系)
National Taiwan University (國立台灣大學)

- 第一章 製造數據科學
- 第二章 製造系統分析與管理
- 第三章 數據科學基礎與模型評估
- 第四章 數據科學分析架構與系統運算決策
- 第五章 數據預處理與製造數據特性
- 第六章 線性分類器
- 第七章 無母數迴歸與分類
- 第八章 決策樹與集成學習
- 第九章 特徵挑選與維度縮減
- 第十章 類神經網路與深度學習
- 第十一章 集群分析
- 第十二章 特徵工程、數據增強與數據平衡
- 第十三章 故障預測與健康管理
- 第十四章 可解釋人工智慧
- 第十五章 概念漂移
- 第十六章 元啟發式演算法
- 第十七章 強化學習

藍：老師課堂講授
綠：學生自學

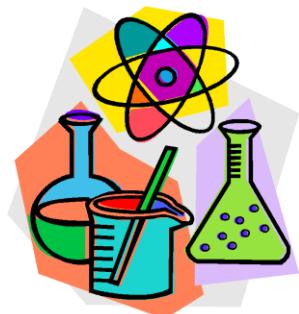
- 附錄A 線性迴歸
- 附錄B 支持向量機
- 附錄C 統計製程管制與先進製程控制
- 附錄D 超參數最佳化
- 應用涵蓋
產能規劃、瑕疵檢測、製程監控與診斷、機台保養、需求預測、生產排程、電腦視覺、自動光學檢測、原料價格預測與採購等

□ This is the course for “Methodology (方法論)”

- Science

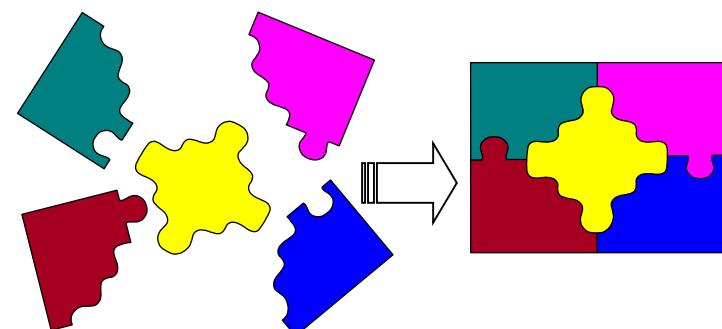
- Emphasize on “**discovery**”
 - impact: long & significant, but 短期可能不見效

- Newton, 1687, *Philosophiæ Naturalis Principia Mathematica*



- Methodology

- Emphasize on “**application**”
 - Impact: The Science of Better (INFORMS)



- Technology

- Emphasize on “**realization/ implementation**”
 - impact: short, but 短期見效
 - LED light: 1965
 - Cell phone: Apr. 3, 1973
 - MS-DOS 1.0: August 1981

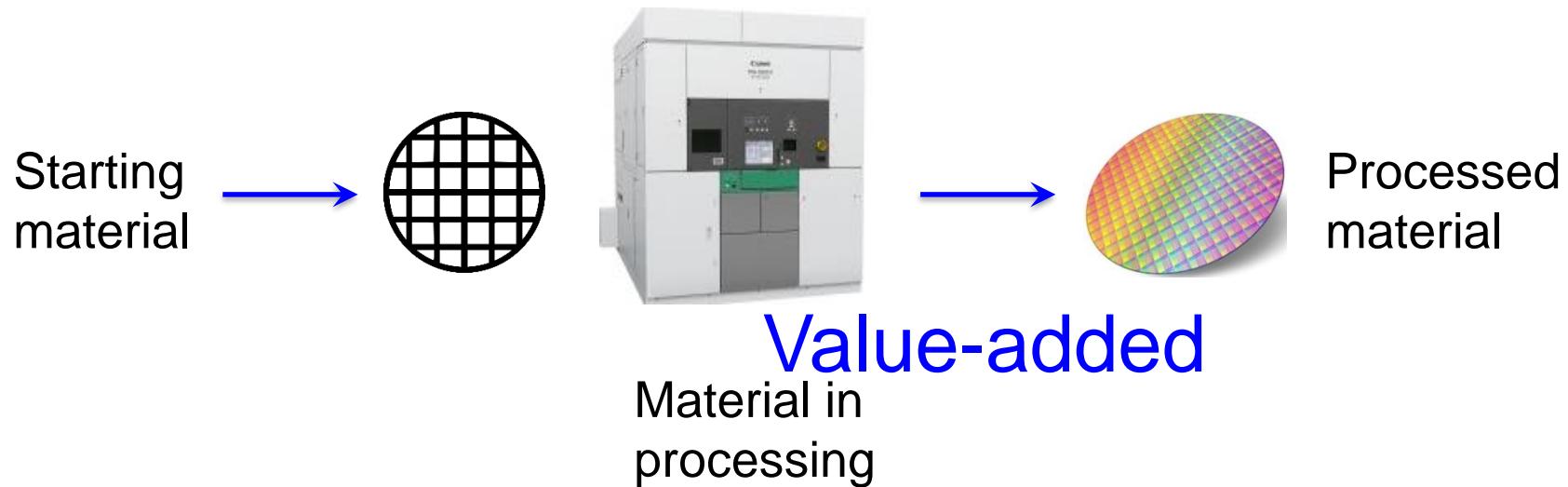


□ What kind of person you want to be

- Scientist
 - Explore the “**truth**” in the world
- Engineer
 - Solve the “**real**” engineering problem based on “**science**”
- Manager
 - Decision making for “**resource**” optimization to achieve the “**goal**”
- Economist
 - Tell a “**story**” to illustrate the “**trading behavior**” among humans
- Lawyer
 - Judge the “**issue**” based on the “**law**”
- Doctor
 - Suggest the “**treatment**” according to the “**root cause**” of “**symptom**”

□ What is “Manufacturing”?

Manus (hand) Factus (make)



Manufacturing is the realization (具現化) of product.

DATA



詮釋的觀點

Information



思辨的過程

Knowledge



科學的原理

修正 (Corrected)

文字化 (Contextualized)

量化 (Quantified)

排序 (Sorting)

計算 (Calculated)

分類 (Categorized)

摘要 (Summarized)

...

重點在於資料品質與資訊詮釋

附加價值 (Value-added):

- Observation
- Interpretation
- Understanding
- Experience
- Skills
- 人:
- Conservation
- Comparison
- Association
- Criticize
- Imagination

重點在於人的參與思辨與討論

迷思：資料？資訊？

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	lot	locn1	locn2	locn3	locn4	locn5	locn6	locn7	locn8	locn9	locn10	locn11	locn12
2	3699	2	3	9	4	1	16	5	10	5	2	6	2
3	1427	9	5	8	3	1		2	10	2	1	16	5
4		2	2	7	1	2	7		5		3	8	
5	3553	9	3	11		4	7		3	9	4	8	4
6	3591	9		6		5	13		10	3		17	
7	3190	9	1	2	3	1	4	5	6			11	2
8	716	1	2	9		2	12	1	3	8	4	3	5
9	91	2	2	5		4	8	3	2	5	3	7	1
10	214	1	1	3	3	3	7	4		2	3	10	1
11	3648	8		6	4	3	5	5		6	4	3	2
12	2128	7	3	5	4	3	11	5		3	2	8	5
13	151		3	8	1	2	3	2	7	2	3	10	2
14	800	6	2	2		5	18	1		2	2	17	1
15	224	7	1		3	4	15	4	2		2	4	5
16	588	4	1	2	5	5	5	5	6	7	3	13	4
17	3798	7	4	2	2	3		4	6	3	3	1	5
18	2012	9	3	3	2	4		1	1	3	4	10	1
19	2895		5	8	4	2	9	4	3		2	4	3
20	1740	3	4	2	1		13	1	7	7	3	18	3
21	3951	5	1		3	2	3	3	11	2	1	19	1
22	3111	6	2	2	4	5	16	2	4	7	3	17	5
23	150	7	1	9	3	1	19	4	4	7	2	11	4
24	commonalityx4000												

Advanced Analytics – Intel: SETFI: Manufacturing data: Semiconductor tool fault isolation. Causality Workbench Repository, <http://www.causality.inf.ethz.ch/repository.php> (2008)

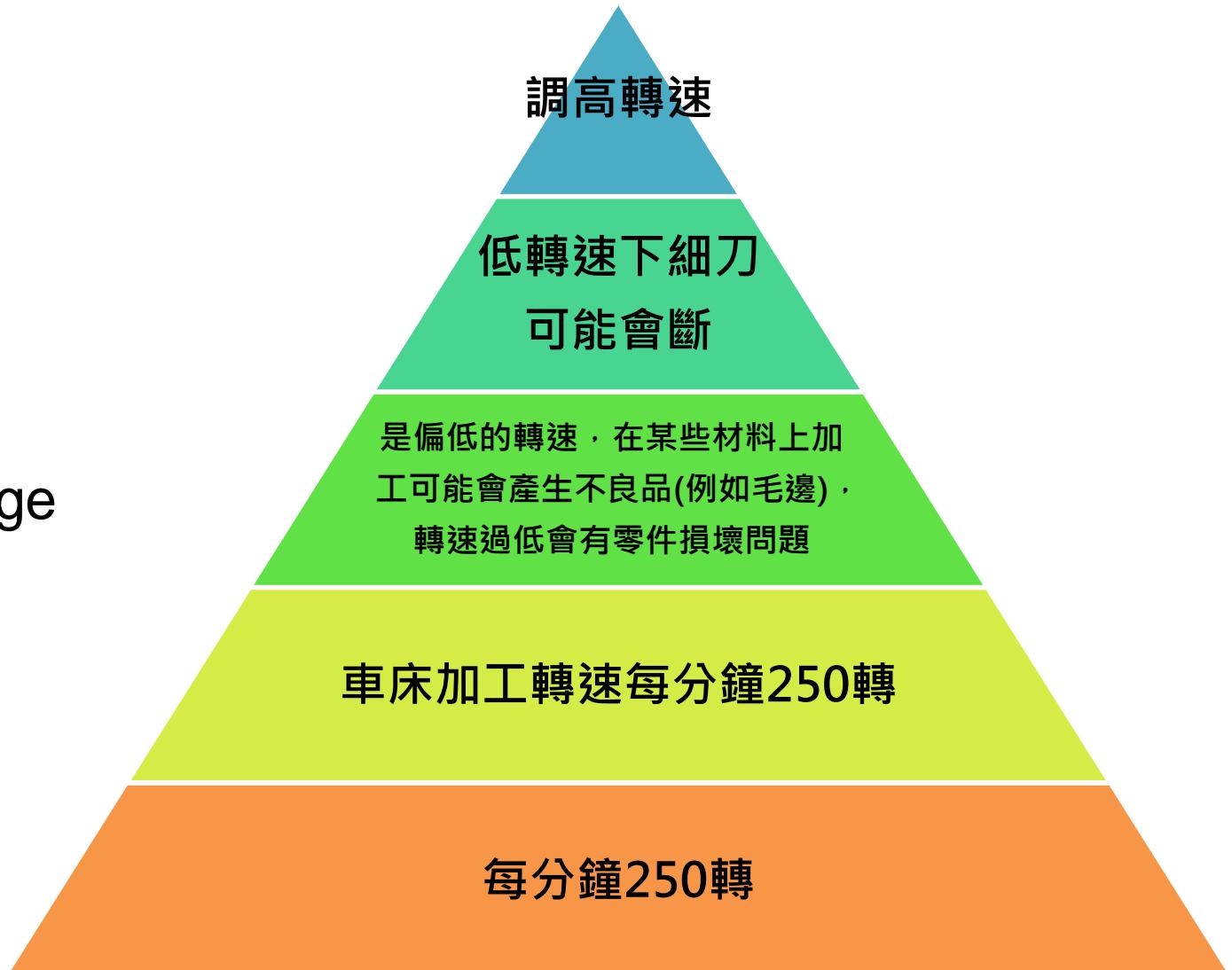
Decision

Insight

Domain Knowledge

Information

Data

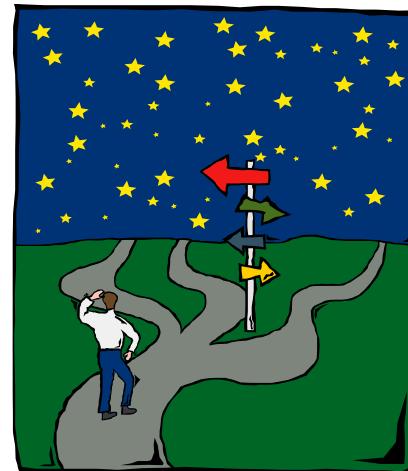


Problem



權衡的觀點

Decision



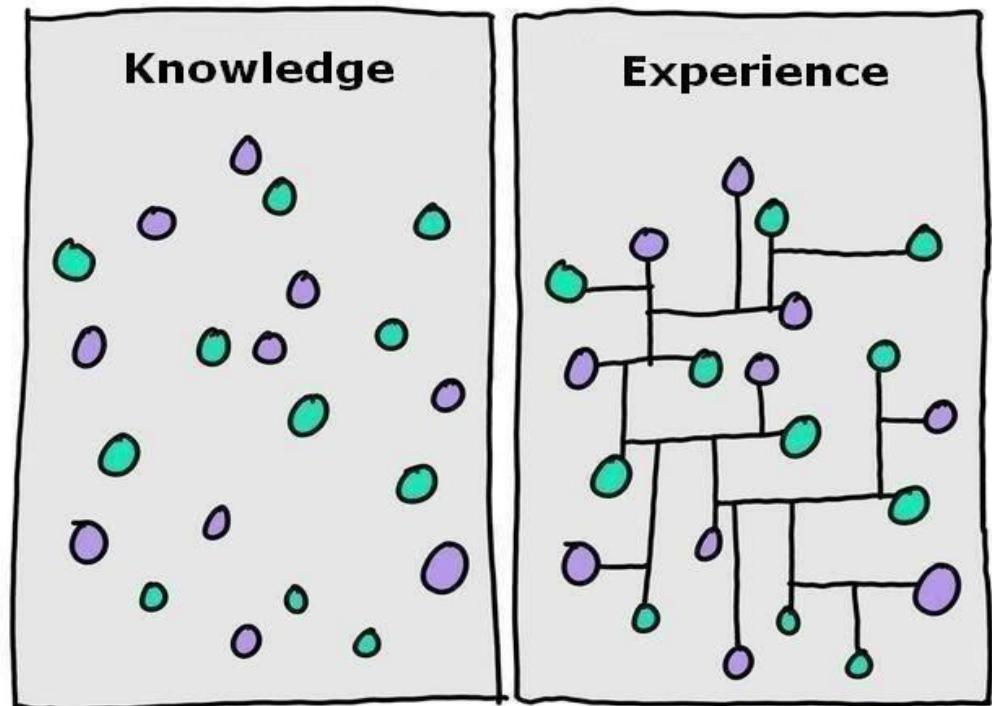
失敗的藝術

Experience



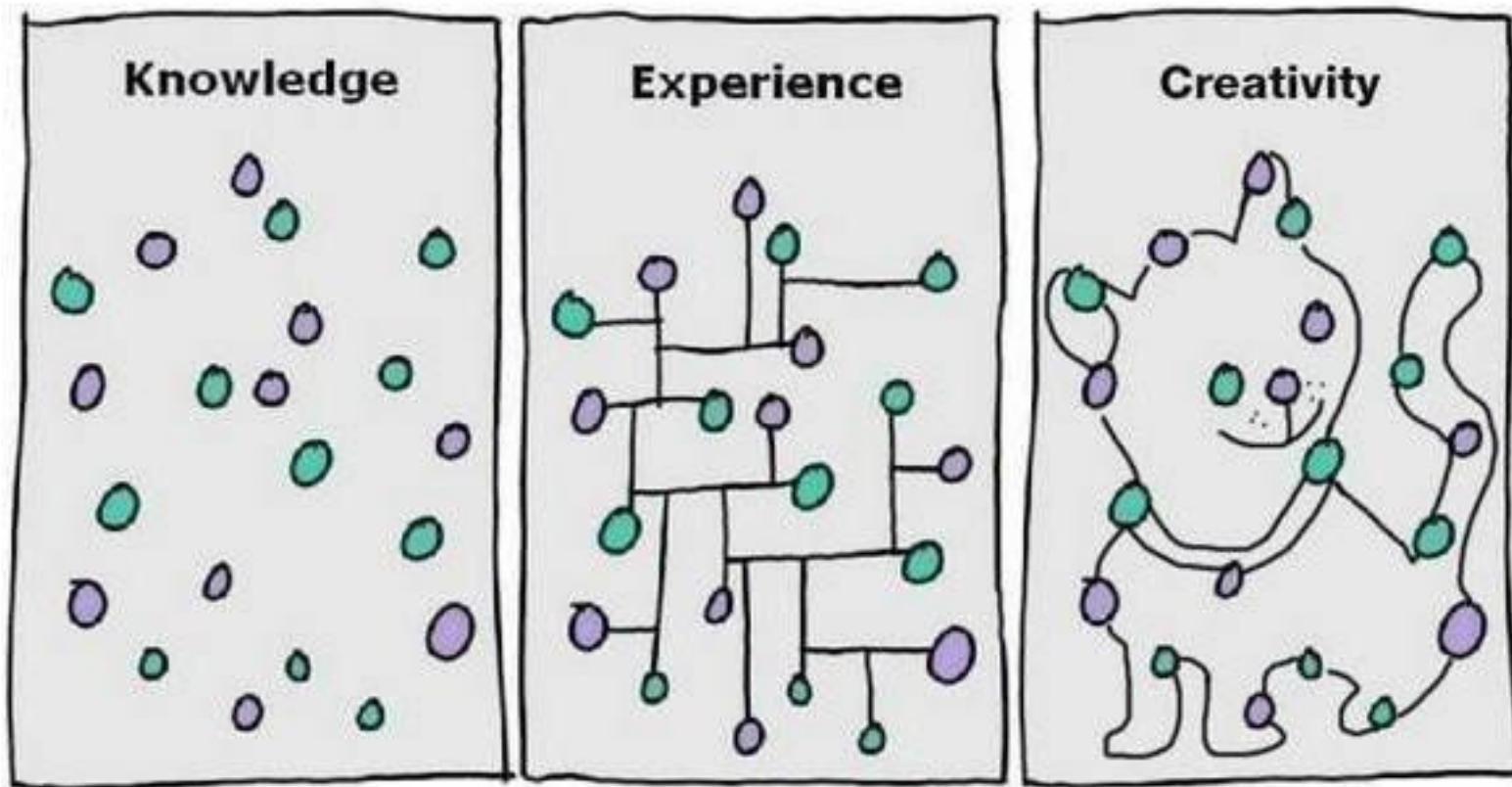
人生的歷練

知識 與 經驗 之 融會貫通 (思辨與討論)



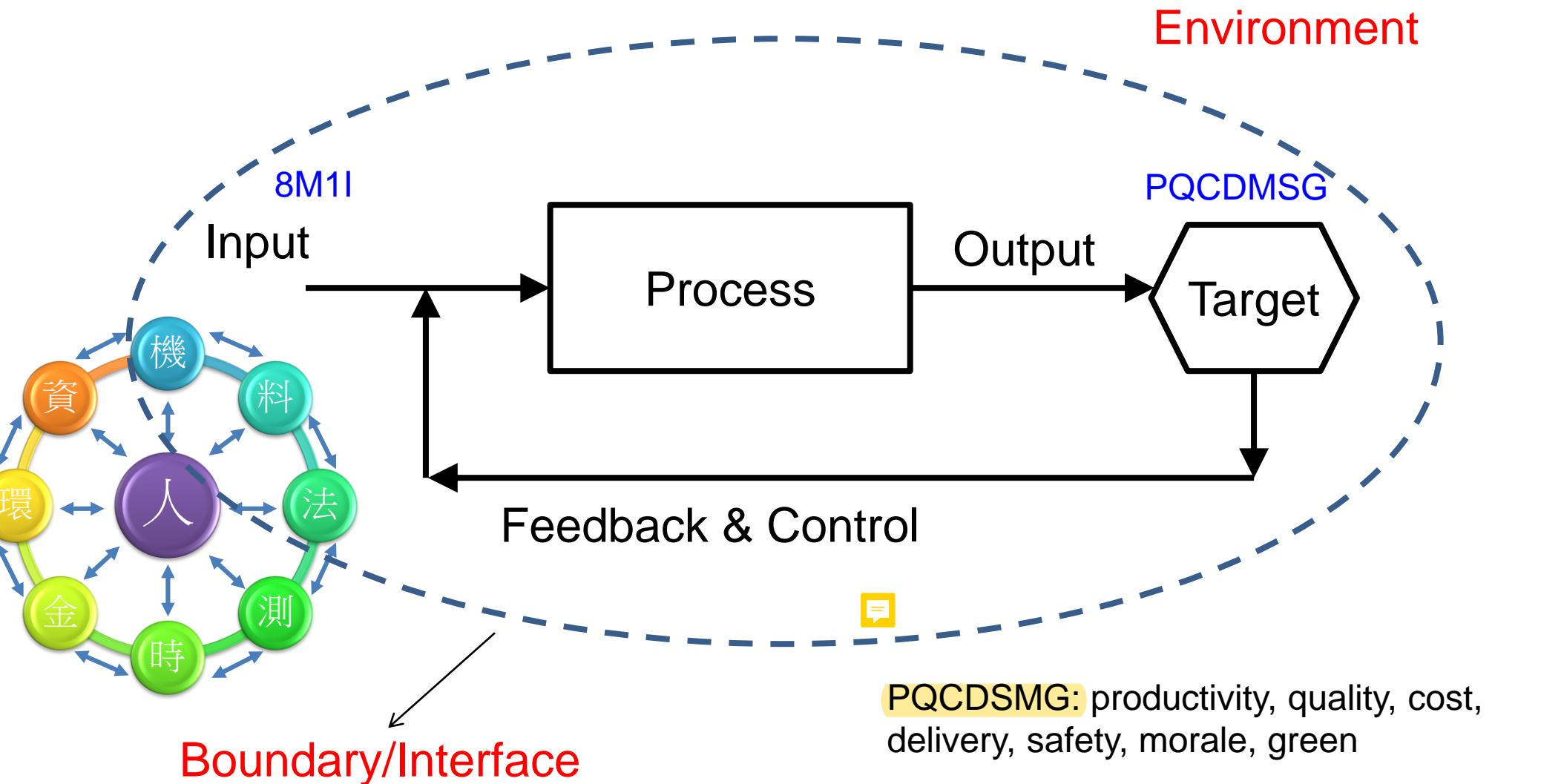
<http://www.herogamingjobs.com/2014/01/07/experience-vs-knowledge/>

- The most important skill in the future will be the ability to “connect the dots” in your own way! (Moioli, 2019)

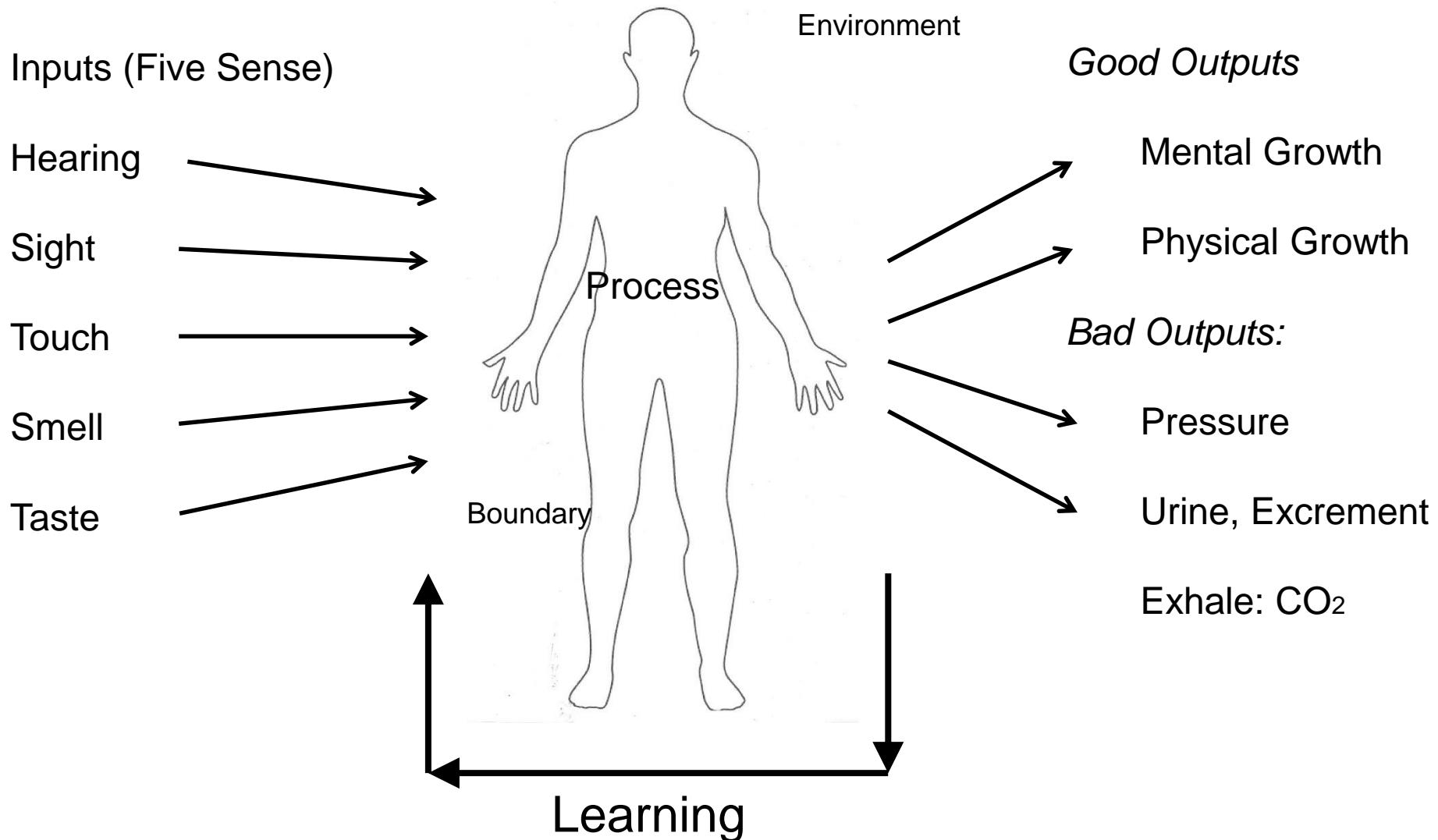


Moioli, Fabio (2019). The most important skill in the future will be the ability to "connect the dots" in your own way!.
<https://www.linkedin.com/pulse/most-important-skill-future-ability-connect-dots-your-fabio-moioli>

□ What is “Systems”?



□ Human Body System



Smart factory is a **decision-oriented** system which has the computational intelligence and self-learning ability to optimize the manufacturing process.

計算智慧 → Based on Data (資料處理與分析)

自我學習 → Real-time Feedback Control (回饋控制)



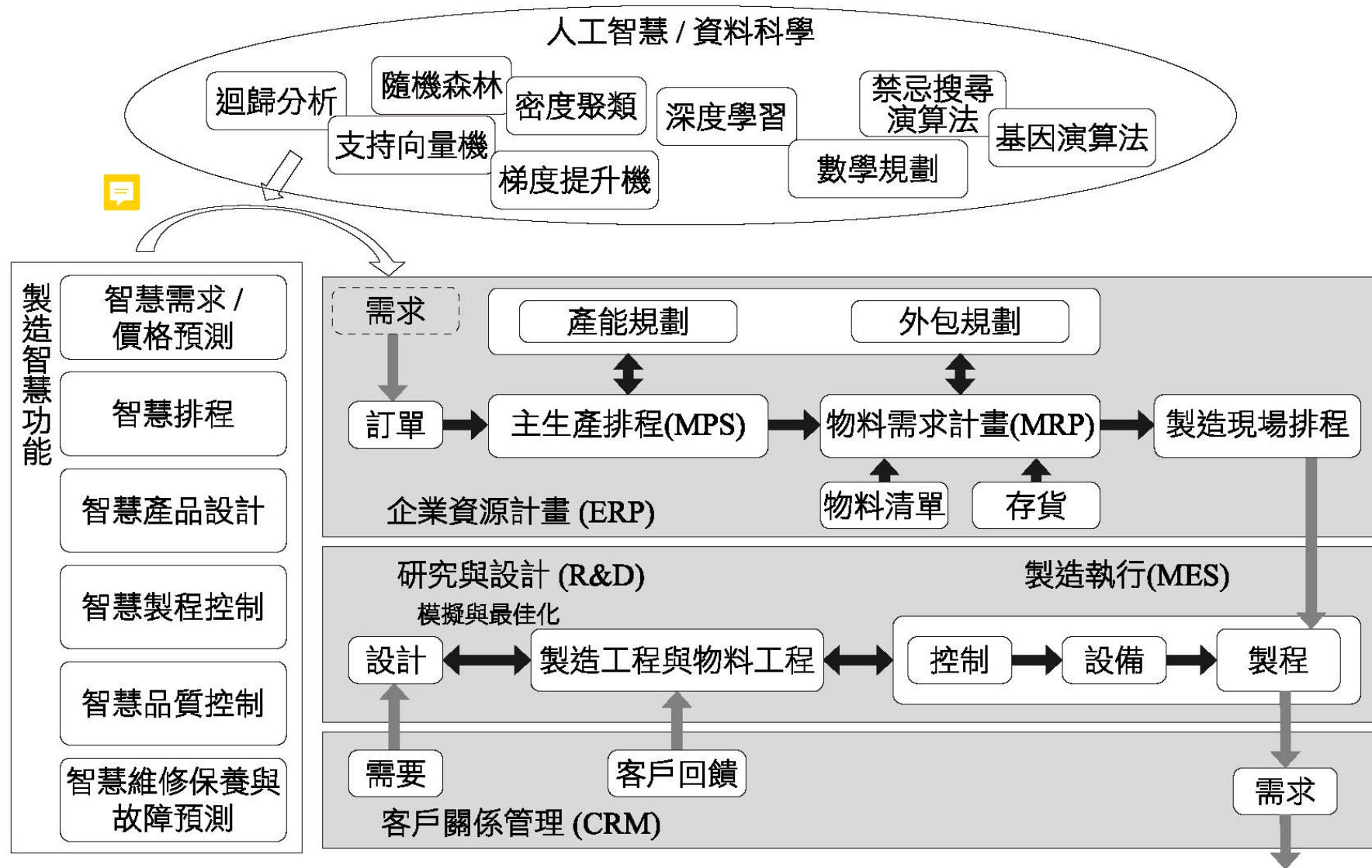
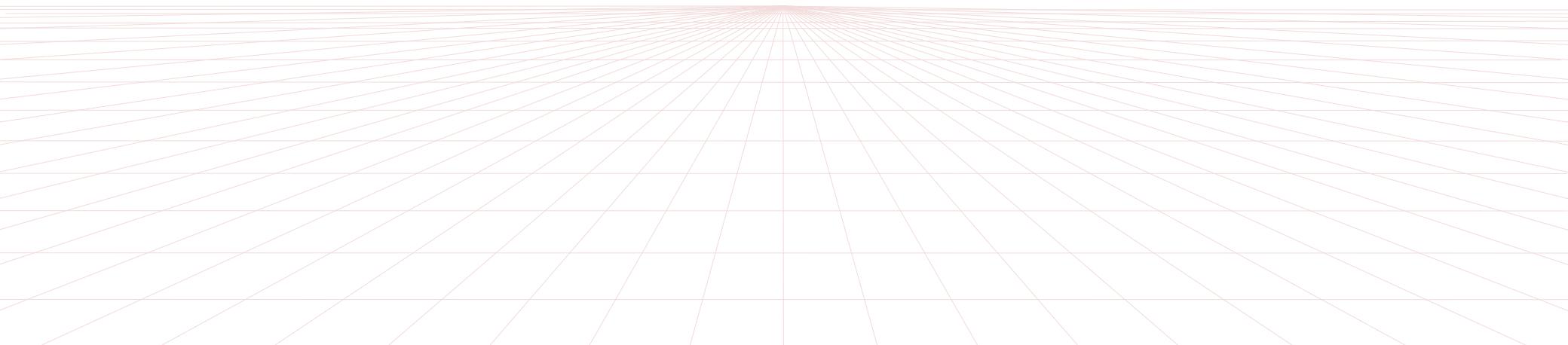


圖 1.5 智慧工廠 (revised from Rao et al., 1993)

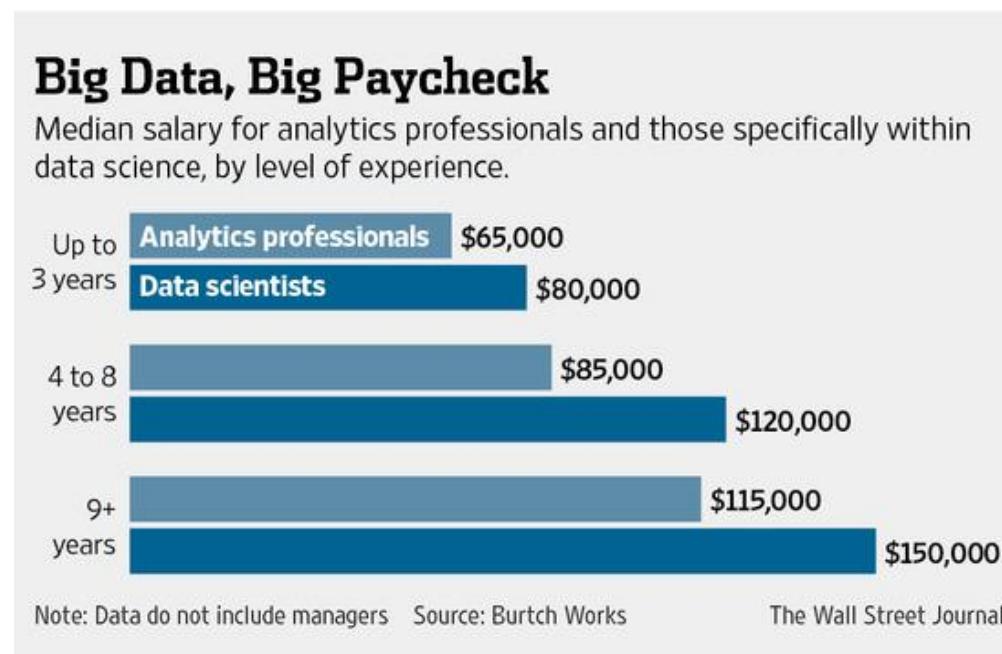
Concluding Remarks



The Sexiest Job in the 21st Century...?

□ "Data Scientist" by Yahoo News 4/4/2013

- "PEHub reports pay for data scientists is upwards of **US\$225,000** even for people straight out of graduate school, up from \$125,000 just a few years ago. For someone with a few years of experience working in the field, pay can reach much higher. One Seattle software-company CEO describes candidates with these skillsets “almost like unicorns.” One got away from the executive when Microsoft approached the data scientist with a **\$650,000** salary plus bonuses.“



<http://finance.yahoo.com/blogs/daily-ticker/sexiest-job-21st-century-122238562.html>

50 Best Jobs in America for 2020



Job Title	Median Base Salary	Job Satisfaction	Job Openings	
#1 Front End Engineer	\$105,240	3.9/5	13,122	View Jobs
#2 Java Developer	\$83,589	3.9/5	16,136	View Jobs
#3 Data Scientist	\$107,801	4.0/5	6,542	View Jobs
#4 Product Manager	\$117,713	3.8/5	12,173	View Jobs
#5 DevOps Engineer	\$107,310	3.9/5	6,603	View Jobs
#6 Data Engineer	\$102,472	3.9/5	6,941	View Jobs
#7 Software Engineer	\$105,563	3.6/5	50,438	View Jobs
#8 Speech Language Pathologist	\$71,867	3.8/5	29,167	View Jobs
#9 Strategy Manager	\$133,067	4.3/5	3,515	View Jobs
#10 Business Development Manager	\$78,480	4.0/5	6,560	View Jobs

Glassdoor, 2020. https://www.glassdoor.com>List/Best-Jobs-in-America-LST_KQ0,20.htm

50 Best Jobs in America for 2021



#1	Java Developer	\$90,830	4.2/5	10,103	View Jobs
#2	Data Scientist	\$113,736	4.1/5	5,971	View Jobs
#3	Product Manager	\$121,107	3.9/5	14,515	View Jobs
#4	Enterprise Architect	\$131,361	4.0/5	10,069	View Jobs
#5	Devops Engineer	\$110,003	4.0/5	6,904	View Jobs
#6	Information Security Engineer	\$110,000	4.0/5	5,621	View Jobs
#7	Business Development Manager	\$82,182	4.1/5	8,827	View Jobs
#8	Mobile Engineer	\$94,301	4.1/5	4,631	View Jobs
#9	Software Engineer	\$110,245	3.8/5	40,564	View Jobs
#10	Dentist	\$134,122	4.0/5	4,315	View Jobs

Glassdoor, 2021. https://www.glassdoor.com>List/Best-Jobs-in-America-LST_KQ0,20.htm

“Data scientists are the people who understand how to fish out answers to important business questions from today’s tsunami of unstructured information.”

Davenport and Patil (2012)

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who
can coax treasure out of
messy, unstructured data.

by Thomas H. Davenport
and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

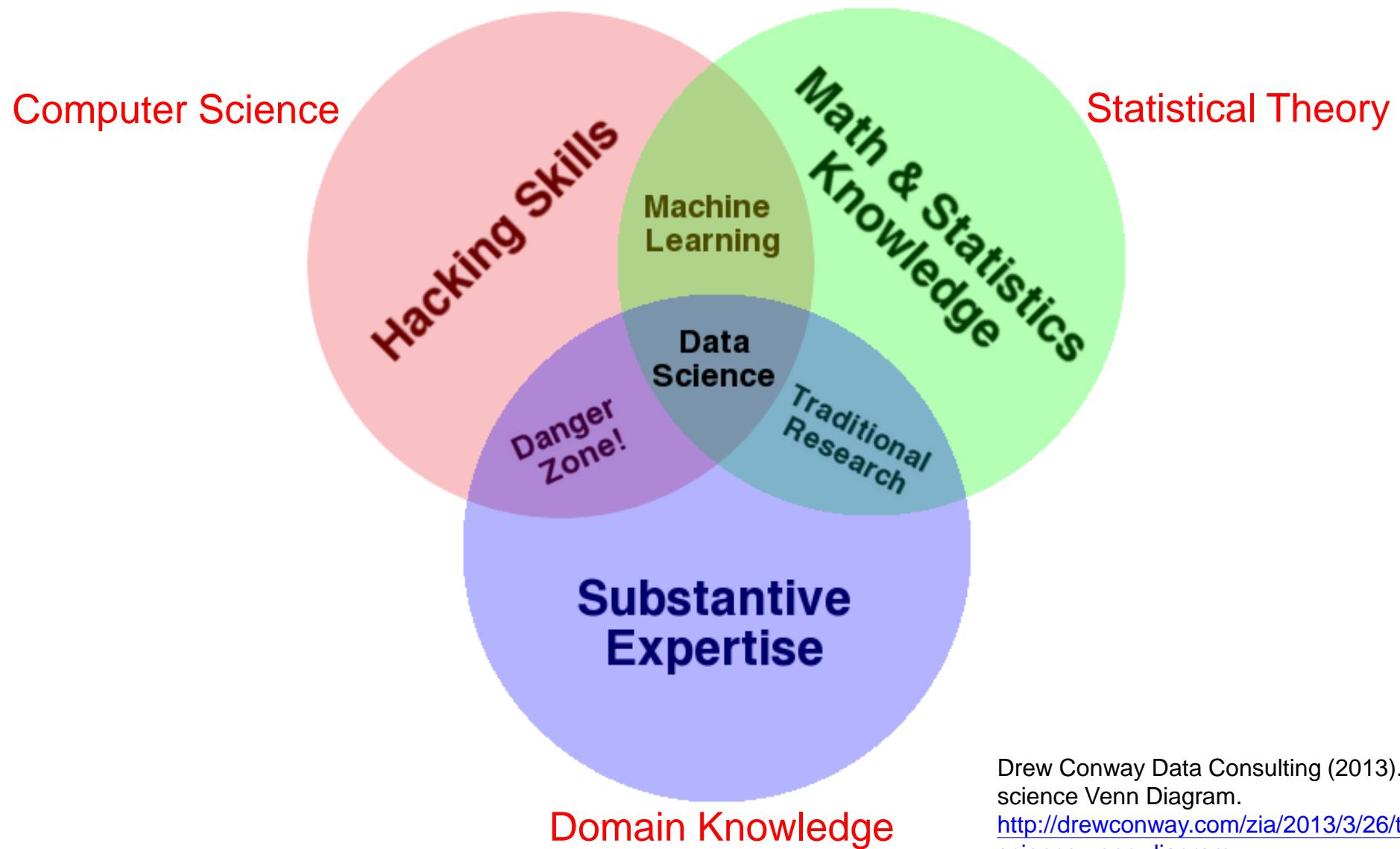
70 Harvard Business Review October 2012



<http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

□ Highlights from HBR

- "...As companies rush to **capitalize** on the potential of big data, the largest constraint many face is the **scarcity of this special talent.**"
- "...The shortage of data scientists is becoming a serious constraint in some sectors."
- "...Data scientists need autonomy but want to be "on the bridge," responding to **management issues** with their managerial colleagues in real time."
- "...Data scientists want to **build things, not just give advice**. One describes **being a consultant** as "the dead zone."
- "...Data scientists today are akin to the Wall Street "quants" of the 1980s and 1990s."



Drew Conway Data Consulting (2013). The data science Venn Diagram.
<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

□ Necessary Skills for Candidates...

- "...user groups devoted to data science tools. The R User Groups (for an open-source **statistical** tool favored by data scientists) and Python Interest Groups (for PIGgies) ..."
- "...hang out with data scientists at the Strata, Structure: Data, and Hadoop World conferences and similar gatherings..."
- "... host a competition on Kaggle or TopCoder, the analytics and coding competition. Follow up with the most-creative entrants..."
- "... **Coding skills** don't have to be at a world-class level but should be good enough to get by. **Look for evidence, too, that candidates learn rapidly about new technologies and methods...**"

□ Necessary Skills for Candidates...

- "...Make sure a candidate can find a **story** in a data set and provide a coherent narrative about a key **data insight**..."
- "...Be wary of candidates who are too detached from the **business world**. When you ask how their work might apply to your management challenges, are they stuck for answers?"
- "...Ask candidates about their favorite analysis or insight and how they are **keeping their skills sharp**. Have they gotten a certificate in the advanced track of Stanford's online Machine Learning course, contributed to open-source projects, or built an online repository of code to share (for example, on GitHub)?"

事實上，製造資料科學是...

釐清問題「定位」
(及影響問題本質
的重要因子)

敘述/診斷性
構析

決策見樹
不見林

預測性構析

掌握環境「趨勢」
(及重要影響因
子的趨勢)

處方性分析補足了
製造資料科學的三大議題

1. 決策風險的測量與權衡
2. 一部份未解決的問題 (eg.
排程、庫存、專案管理)
3. “槓桿”有限資源下的最佳化
(人、機、料、法、環)

處方/自動性
構析

MDS

問題本質
欠缺瞭解



建立決策「格局」
(資源佈署)

	智慧製造系統（IMS）	製造數據科學（MDS）
本質	系統開發	分析架構
議題	機器學習技術解決問題	問題本質探索與整體系統性影響
觀點	演算法與技術	資源與流程
數據	流程內的數據	流程內與流程間數據
點線面	問題點	問題面
問題解決	機台效率或產品良率等問題	跨部門跨流程資源整合等問題 (化解跨組織間的目標衝突)
整合	資通訊技術（ICT）	資通訊技術（ICT）與運營技術（OT）
產出	自動化系統或模組	數據科學分析架構與資源優化
工程與管理	著重在工程技術	涉及管理議題
案例或模組	製造執行系統（manufacturing execution system, MES）、系統監控和資料擷取（supervisory control and data acquisition, SCADA）、SPC、APC、VM、錯誤偵測與分類（fault detection and classification, FDC）、自動光學檢測（AOI）、生產排程系統等	企業資源規劃（enterprise resource planning, ERP）、顧客關係管理（customer relationship management, CRM）、產品生命週期管理（product lifecycle management, PLM）、故障預測與健康管理、產品組合優化、冰機加減機能能源控管、產能規劃 - 資本支出 - 成本結構（capacity-capex-cost）等

Road to Data Scientist ...

Dept. of Statistics

Probability & Statistics
Operations Research
Multivariate Analysis

Dept. of Computer Science

Data Structure
Algorithm
Database Management

Production(產)
Marketing(銷)
HR(人)
R&D(發)
Finance(財)
Information(資)

Management
Insight

Graduate Level

Data Mining
Cloud/Edge Computing
Statistical Learning Theory

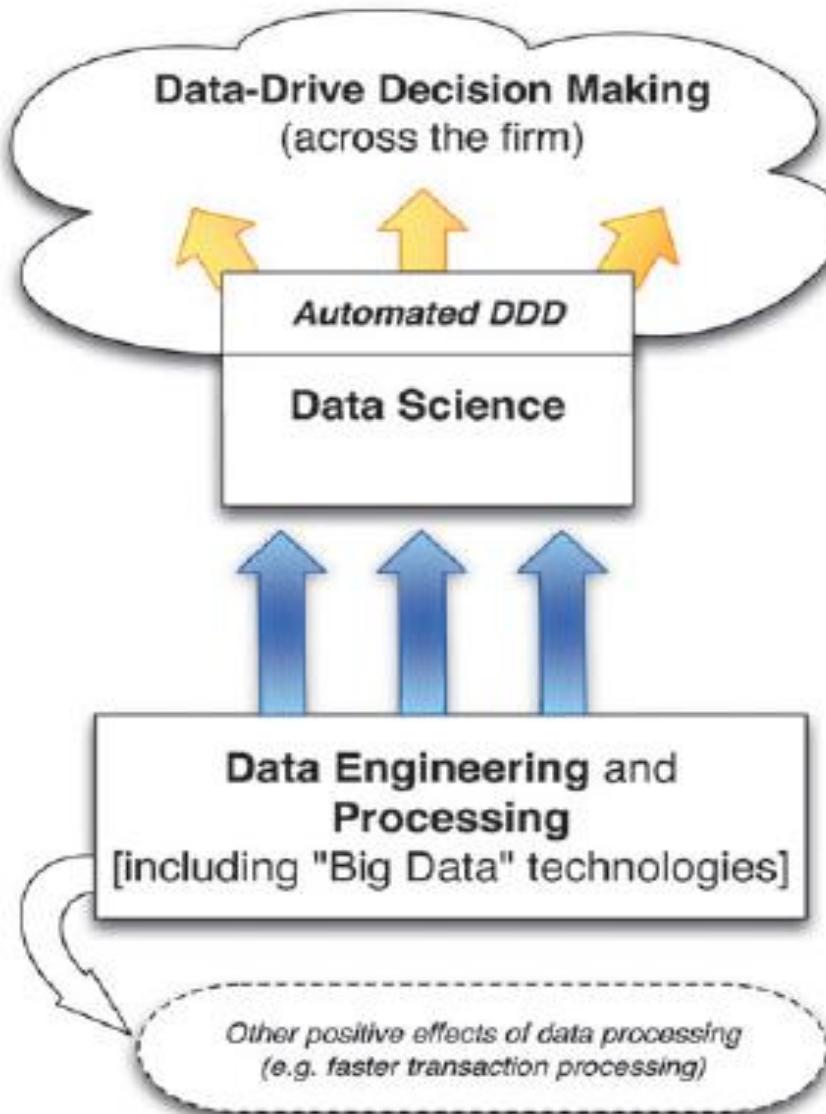
Data-Driven Decision Making

Applications

Signal Processing/
Time Series
Natural Language
Processing/
Text Mining
Domain
Knowledge
Computer Vision/ Image
Processing
Speech/Voice
Processing

Source: Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51-59.

Data Science in the Organization



Source: Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51-59.

□ Kaggle (<https://www.kaggle.com/>)

General InClass Sort by Grouped

All Categories Search competitions

17 Active Competitions

	2018 Data Science Bowl Find the nuclei in divergent images to advance medical discovery <small>Featured · 2 months to go · 🧠 biology</small>	\$100,000 1,583 teams
	Mercari Price Suggestion Challenge Can you automatically suggest product prices to online sellers? <small>Featured · 15 hours to go ·</small>	\$100,000 2,386 teams
	Google Cloud & NCAA® ML Competition 2018-Men's Apply Machine Learning to NCAA® March Madness® <small>Featured · a month to go ·</small>	\$50,000 36 teams
	Google Cloud & NCAA® ML Competition 2018-Women's Apply machine learning to NCAA® March Madness® <small>Featured · a month to go ·</small>	\$50,000 27 teams
	Toxic Comment Classification Challenge	\$35,000

□ UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>)



The screenshot shows the homepage of the UCI Machine Learning Repository. At the top left is the UCI logo with a drawing of an antelope. To the right are links for "About", "Citation Policy", "Donate a Data Set", and "Contact". Below the logo is a search bar with a "Google Custom Search" placeholder and a "Search" button. A link "View ALL Data Sets" is also visible. The main content area features the repository's name and subtitle "Center for Machine Learning and Intelligent Systems".

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 272 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable interface. Our [old web site](#) is still available, for those who prefer the old format. For a general overview of the Repository, please visit our [About page](#). For information about citing data sets in publications, please read our [citation policy](#). If you wish to donate a data set, please consult our [donation policy](#). For any other questions, feel free to [contact the Repository librarians](#). We have also set up a [mirror site](#) for the Repository.



Latest News:

2013-04-04: Welcome to the new Repository admins Kevin Bache and Moshe Lichman!
2010-03-01: [Note](#) from donor regarding Netflix data
2009-10-16: Two new data sets have been added.
2009-09-14: Several data sets have been added.
2008-07-23: [Repository mirror](#) has been set up.
2008-03-24: New data sets have been added!
2007-06-25: Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

Newest Data Sets:

2014-02-12:  HIGGS
2014-02-12:  SUSY
2014-02-05:  EMG dataset in Lower Limb
2014-01-09:  SML2010

Most Popular Data Sets (hits since 2007):

524283:  Iris
367595:  Adult
315661:  Wine
261273:  Breast Cancer Wisconsin (Diagnostic)

□ Kdnuggets

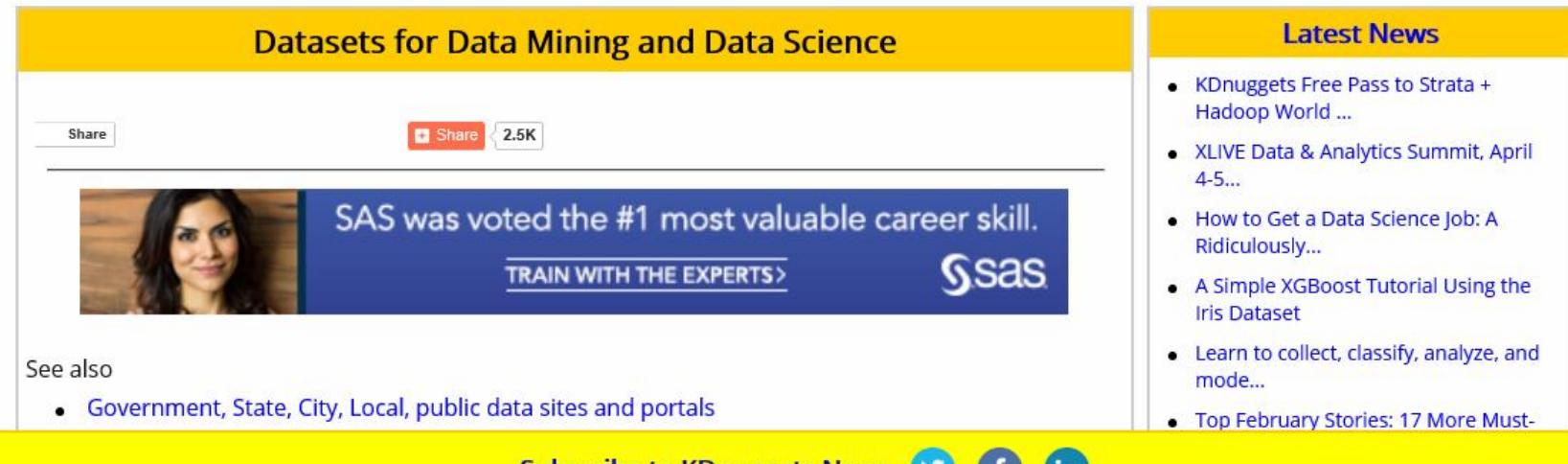
(<http://www.kdnuggets.com/datasets/index.html>)



The banner features the KDNuggets logo, social media links (Twitter, Facebook, LinkedIn), a search bar, and a navigation menu with categories like SOFTWARE, NEWS, Top stories, Opinions, Tutorials, JOBS, Companies, Courses, Datasets, EDUCATION, Certificates, Meetings, and Webinars. Below the menu is a promotional image for the PAW San Francisco conference, showing a speaker on stage and the text "Empower your business with predictive analytics". A "LEARN MORE" button is also present.

PAW San Francisco, May 14-18: Empower your business with Predictive Analytics. Register today!

[KDNuggets Home](#) » [Datasets](#)



The page title is "Datasets for Data Mining and Data Science". It includes a "Share" button, a "2.5K" share count, and a featured ad for SAS. The ad shows a woman's portrait and the text "SAS was voted the #1 most valuable career skill. TRAIN WITH THE EXPERTS! sas". On the right side, there is a "Latest News" sidebar with a list of articles:

- KDNuggets Free Pass to Strata + Hadoop World ...
- XLIVE Data & Analytics Summit, April 4-5...
- How to Get a Data Science Job: A Ridiculously...
- A Simple XGBoost Tutorial Using the Iris Dataset
- Learn to collect, classify, analyze, and mode...
- Top February Stories: 17 More Must-

Below the main content, there is a "See also" section with a link to "Government, State, City, Local, public data sites and portals". At the bottom, there is a "Subscribe to KDNuggets News" button and social media links.

□ 政府資料開放平台 (<https://data.gov.tw/>)



The screenshot shows the main navigation bar with links for '全部資料集', '互動專區', '最新消息', '諮詢小組', '關於平臺', and 'ENGLISH'. It also features a search bar and a banner stating '還在爬蟲嗎？本平臺35192筆資料集全都放在這裡'.

資料集服務分類



生育保健(344)



出生及收養(46)



求學及進修(561)



服兵役(181)



求職及就業(498)



開創事業(427)



婚姻(4)



投資理財(1532)



休閒旅遊(844)



交通及通訊(1588)



就醫(845)



購屋及遷徙(577)

□ <https://www.phmsociety.org/competition/phm/10>



Username or e-mail: *

Password: *

[Log in](#)

New to the PHM Society? [Create an account](#). Forgot your password? [Request a new password](#).

[home](#) [about](#) [conferences](#) [journal](#) [competition](#) [publications](#) [community](#) [sponsors](#) [partners](#) [awards](#)



[Home](#) » 2010 PHM Society Conference Data Challenge

2010 phm society conference data challenge

The PHM Data Challenge is a competition open to all potential conference attendees. This year the challenge is focused on RUL estimation for a high-speed CNC milling machine cutters using dynamometer, accelerometer, and acoustic emission data.

Both Student and Professional teams are encouraged to enter! Winners of the Student and the Professional categories who attend the conference and submit an invited paper to ijPHM on their technique will be awarded a cash prize. Top scoring participants will be invited to present at a special session of the conference.

Participants will be scored based on their ability to estimate the remaining useful life of a 6mm ball nose tungsten carbide cutter. Winners of the Student and the Professional categories who attend the conference and submit an invited paper to ijPHM on their technique will be awarded a cash prize. Top scoring participants will be invited to present at a special session of the conference.

Additional information can be found on the competition blog, <http://www.phmsociety.org/forum/583>

Teams

Teams may be comprised of one or more researchers. One winner from each of two categories will be determined on the basis of score. The categories are:

quick links

[Journal](#)

[Forum](#)

[User Directory](#)

[PHM Jobs](#)

[Submit Conf. Paper \(Instructions\)](#)

[Submit IJPHM Paper](#)

upcoming events

+ [IEEE Aerospace Conference 2018, Big Sky, MT from 3 - 10 March 2018](#)

+ [MFPT 2018 Conference, Virginia Beach, VA from 15 - 17 May 2018](#)

+ [European Conference of the PHM Society 2018 \(PHME18\) Utrecht, The Netherlands from 3 - 6 July 2018](#)

+ [Annual Conference of the PHM Society 2018 \(PHM18\) Philadelphia, PA from 24 - 27 September 2018](#)

+ [Related Conferences \(via MFPT\)](#)

□ 在「製造數據科學」的思維中，除了工程技術面的自動化，也涉及管理或組織的相關議題。

- 「策略」(strategy)是「重點」的選擇 (司徒達賢，2016)；而「管理」(management)是「群策群力，以竟事功。」(許士軍，2019)，使目標或事情圓滿達成的活動過程。
- 管理本質上就是「管資源、理流程」；可簡化為「管好人、理好事」，在從管理資源 (8M1I) 的活動中，來達成績效目標 (PQCDSMG) 。

資訊管理

倚資訊來做**資源與流程**上的槓桿

- 金屬扣件智慧製造系統關鍵技術 - 金屬中心
 - <https://www.youtube.com/watch?v=KbQEhx6Ky8Q>



Thanks for your attention



NTU Dept. of Information Management
name: 李家岩 (FB: Chia-Yen Lee)
phone: 886-2-33661206
email: chiayenlee@ntu.edu.tw
web: <https://polab.im.ntu.edu.tw/>