وكلاء الذكاء الاصطناعي في المعلوماتية الحيوية

Agentic AI in Bioinformatics

دليل عملي لنشر الوكلاء الأذكياء في المهام اليومية للباحث

د. فضل محمد الأكوع

Dr. Fadhl Mohammed Alakwaa

Arabic & English | عربي - إنجليزي

2026

وكلاء الذكاء الاصطناعي في المعلوماتية الحيوية

Agentic AI in Bioinformatics

دليل عملي لنشر الوكلاء الأذكياء في المهام اليومية للباحث

د. فضل محمد الأكوع

Dr. Fadhl Mohammed Alakwaa

الطبعة الأولى — 2026

جميع الحقوق محفوظة للمؤلف

أُنتج هذا الكتاب بمساعدة نموذج الذكاء الاصطناعي Claude Opus 4.8 من شركة Anthropic.

This book was produced with the assistance of Claude Opus 4.8 (Anthropic).

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

إهداء

إلى كلِّ باحثٍ في علوم الحياة يقضي لياليَه أمام الشاشة يفكُّ شيفرةَ الجينات، ويلاحق سرَّ الخلية، ويبحث في بحارٍ من البيانات عن قبسٍ من المعرفة ينفع الناس. إلى المعلوماتيِّ الحيوي الذي يجمع بين دقة البيولوجيا وصرامة الحوسبة، وإلى من يؤمنون بأن تسخير الذكاء الاصطناعي لخدمة الطب والصحة أمانةٌ علميةٌ قبل أن يكون إنجازًا تقنيًّا. إلى زملائي وطلابي في مختبرات البحث، وإلى كلِّ يدٍ تسعى لأن تكون أدواتُ الغد صُنعَ أمتنا لا مجردَ مستهلَكٍ لها.

To every researcher in the life sciences who spends their nights before the screen decoding genes, chasing the secret of the cell, and searching seas of data for a spark of knowledge that benefits humankind. To the bioinformatician who combines the precision of biology with the rigor of computation, and to those who believe that harnessing artificial intelligence in the service of medicine and health is a scientific trust before it is a technical achievement. To my colleagues and students in the research laboratories, and to every hand striving to make tomorrow's tools the making of our nation, not merely something it consumes.

المحتويات

المقدمة

٠١ لماذا الوكلاء الأذكياء في المعلوماتية الحيوية؟

٠٢ خريطة المهام اليومية وأين تنشر الوكلاء

٠٣ تجهيز بيئة العمل الوكيلية

٠٤ وكلاء مراجعة الأدبيات وتركيب المعرفة

٠٥ استرجاع البيانات والاستعلام عن قواعد البيانات

٠٦ وكلاء تحليل التسلسلات

٠٧ تنسيق خطوط الأنابيب وأتمتة سير العمل

٠٨ معالجة البيانات والتحليل الإحصائي

٠٩ توليد الشيفرة والتصحيح وقابلية إعادة الإنتاج

١٠. وكلاء التصوير وإعداد التقارير

١١. أنظمة الوكلاء المتعددة للتحليلات المعقدة

١٢. التحقق والأمان والصرامة العلمية

١٣. دراسة حالة: تحليل متكامل بالوكلاء

١٤. الأخلاقيات وخصوصية البيانات والمستقبل

الخاتمة

المراجع العلمية

English Section

نبذة عن المؤلف

مؤلفات المؤلف

القسم العربي

Arabic Section

المقدمة

﴿ اقْرَأْ وَرَبُّكَ الْأَكْرَمُ ﴾

العلق: ٣

تعيش المعلوماتيةُ الحيوية لحظةً تحوُّليةً غيرَ مسبوقة. فمنذ أن أصبح تسلسلُ الجينوم البشري متاحًا، انفجرت كميةُ البيانات البيولوجية انفجارًا يفوق قدرةَ أيِّ باحثٍ فردٍ على استيعابها: تسلسلاتٌ جينيةٌ بمليارات القواعد، وتجاربُ تعبيرٍ جينيٍّ لآلاف الخلايا، وقواعدُ بياناتٍ تنمو كلَّ يوم. وفي خضمِّ هذا الطوفان، يقضي المعلوماتيُّ الحيوي معظمَ يومه في مهامَّ متكررةٍ ومتشعبة: البحث في الأدبيات، واستخراج البيانات من قواعدها، وكتابة سكربتاتٍ لمعالجة الأرقام، وتصحيح أخطاءٍ برمجية، وإعداد التقارير. هنا بالضبط يأتي دورُ الوكلاء الأذكياء.

هذا الكتاب ليس مقدمةً نظريةً عامة في الذكاء الاصطناعي، بل دليلٌ عمليٌّ موجَّهٌ لمن يعمل في المعلوماتية الحيوية فعلًا. غايتُه أن يجيب عن سؤالٍ محدد: كيف أنشر الوكلاء الأذكياء في مهامي اليومية لأعمل أسرعَ وأدقَّ دون أن أفرِّط في الصرامة العلمية؟ سننتقل من المبدأ إلى التطبيق المباشر: من البحث في قواعد بيانات NCBI و

Ensembl و UniProt، إلى تحليل تسلسلات الحمض النووي، إلى تنسيق خطوط أنابيب التحليل، إلى توليد الشيفرة البرمجية وتصحيحها، إلى التصوير وإعداد التقارير.

لكنَّ هذا الكتاب يحمل تحذيرًا بقدر ما يحمل وعدًا. فالبيانات البيولوجيةُ حساسةٌ، وبعضُها بياناتٌ صحيةٌ خاصةٌ بمرضى يحميها القانون. والاستنتاجُ الخاطئ في علمٍ يمسُّ صحةَ الإنسان قد تكون له عواقبُ وخيمة. لذلك فإن نشر الوكلاء في البحث العلمي يتطلب حذرًا مضاعفًا: التحقُّقُ من كل مخرج، وحفظُ خصوصية البيانات، وضمانُ قابلية إعادة إنتاج النتائج. الوكيلُ في المعلوماتية الحيوية مساعدٌ قويٌّ لا بديلٌ عن حكم الباحث ومسؤوليته.

كتبتُ هذا الكتاب من موقع الممارس، باحثًا في البيولوجيا الحاسوبية يستخدم هذه الأدواتِ في عمله اليومي. وحرصتُ أن يكون كلُّ فصلٍ مرتبطًا بمهمةٍ حقيقيةٍ يواجهها القارئ، لا بتجريدٍ بعيد. وجعلتُ كلَّ فصلٍ يفتح بآيةٍ كريمةٍ تذكّرنا بأن دراسةَ الحياة في خلاياها وجيناتها هي في جوهرها تأملٌ في بديع خلق الله، وأن العلم النافع عبادةٌ ومسؤولية.

أدعوك أن تقرأ هذا الكتاب وأنت أمام شاشتك، تجرِّب ما تقرؤه على بياناتك ومهامك. فالإتقانُ يأتي بالممارسة، والوكيلُ الذي تبنيه لمهمتك أنت أنفعُ من ألف مثالٍ نظري. ولنبدأ على بركة الله.

الفصل ١

لماذا الوكلاء الأذكياء في المعلوماتية الحيوية؟

﴿ وَفِي أَنفُسِكُمْ ۚ أَفَلَا تُبْصِرُونَ ﴾

الذاريات: ٢١

المعلوماتيةُ الحيوية حقلٌ يقع عند تقاطع البيولوجيا والحوسبة والإحصاء، وهو من أكثر الحقول استنزافًا للوقت في تفاصيلَ تقنيةٍ متكررة. فالباحثُ النموذجي قد يقضي ساعاتٍ في تنسيق ملفِّ بياناتٍ ليقبله برنامجٌ ما، أو في تذكُّر صيغة أمرٍ في سطر الأوامر، أو في البحث عن السبب وراء رسالة خطأ غامضة. هذه المهامُّ، رغم ضرورتها، لا تتطلب الإبداعَ العلميَّ الذي تدرَّب الباحثُ من أجله. وهنا تكمن الفرصةُ الكبرى للوكلاء الأذكياء.

طبيعة العمل البيولوجي الحاسوبي

يتميز العمل في المعلوماتية الحيوية بسماتٍ تجعله مرشَّحًا مثاليًّا لأتمتةٍ ذكية. أولًا، كثرةُ المهام المتكررة القابلة للوصف بدقة: تنزيلُ بيانات، تحويلُ صيغ، تشغيلُ أدواتٍ بمعاملاتٍ محددة. ثانيًا، الحاجةُ الدائمة إلى الربط بين أدواتٍ وقواعدِ بياناتٍ متعددة،

إذ نادرًا ما يكتمل تحليلٌ بأداةٍ واحدة. ثالثًا، حجمُ المعرفة الهائل المتغير باستمرار: مقالاتٌ جديدة كلَّ يوم، وأدواتٌ تُحدَّث، وقواعدُ بياناتٍ تنمو. هذه السماتُ الثلاث — التكرار، والترابط، وتدفق المعرفة — هي بالضبط ما يبرع فيه الوكلاء.

ما الذي يضيفه الوكيل فوق المساعد البسيط؟

قد يسأل الباحث: ألا يكفيني مساعدٌ لغويٌّ أسأله فيجيبني؟ الفرقُ جوهري. المساعدُ البسيط يقترح عليك أمرًا لتنفِّذه أنت بنفسك، أما الوكيلُ فينفِّذه ويقرأ نتيجته ويبني عليها. تخيَّل الفرق بين أن يقول لك المساعد «جرِّب هذا الأمر للبحث في NCBI» وبين أن يبحث الوكيلُ فعلًا، ويقرأ النتائج، ويصفِّيها، ويلخِّصها لك في تقرير. الوكيلُ يغلق الحلقةَ بين الفكرة والتنفيذ، فيتحول من مستشارٍ إلى منفِّذٍ يعمل تحت إشرافك.

أين يناسب الوكيل وأين لا يناسب؟

ليس كلُّ شيءٍ في البحث صالحًا للأتمتة. الوكلاءُ ممتازون في المهام الواضحة المعالم: استرجاع البيانات، تحويل الصيغ، توليد الشيفرة الروتينية، البحث الأولي في الأدبيات. لكنهم ضعفاء — وخطرون أحيانًا — في القرارات العلمية الجوهرية: اختيارُ الفرضية، وتأويلُ النتائج البيولوجية، والحكمُ على دلالة اكتشافٍ ما. القاعدةُ الذهبية أن تترك للوكيل العملَ الميكانيكيَّ الثقيل، وتحتفظ لنفسك بالحكم العلمي. فالوكيلُ يوفِّر لك الوقتَ والجهد لتركِّز طاقتك حيث تكون قيمتُك الحقيقية بوصفك باحثًا.

الفصل ٢

خريطة المهام اليومية وأين تنشر الوكلاء

﴿ وَجَعَلْنَا مِنَ الْمَاءِ كُلَّ شَيْءٍ حَيٍّ ﴾

الأنبياء: ٣٠

قبل أن ننشر أيَّ وكيل، علينا أن نرسم خريطةً واضحةً ليومنا: ما المهامُّ التي نكررها؟ وكم تستغرق؟ وأيُّها يستنزف وقتًا لا يستحقه؟ هذا التحليلُ البسيط هو حجرُ الأساس، لأن الأتمتةَ الذكية تبدأ بمعرفة ما الذي يستحق الأتمتة. لنستعرض المهامَّ النموذجية ليوم المعلوماتي الحيوي ونحدد أين يضيف الوكيلُ أكبرَ قيمة.

دورة حياة المشروع البحثي

يمر أيُّ مشروعٍ في المعلوماتية الحيوية بمراحلَ متشابهة. يبدأ بمراجعة الأدبيات لفهم ما أُنجز وتحديد الفجوة. ثم يأتي جمعُ البيانات واسترجاعُها من قواعدها العامة أو من تجارب المختبر. تليها مرحلةُ معالجة البيانات وتنظيفها وضبط جودتها، وهي من أكثر المراحل استهلاكًا للوقت. ثم التحليلُ الأساسي والإحصائي، فالتصويرُ وتأويلُ النتائج،

وأخيرًا كتابةُ التقرير أو المقالة العلمية. في كل مرحلةٍ من هذه المراحل ثمة مهامٌّ روتينيةٌ يمكن للوكيل أن يتولاها.

تصنيف المهام حسب قابلية الأتمتة

يمكن تصنيفُ المهام في ثلاث فئات. الفئةُ الأولى: مهامٌّ آمنةُ الأتمتة بالكامل، كتنزيل ملفٍ من قاعدة بيانات، أو تحويل صيغة، أو توليد رسمٍ بياني قياسي؛ هذه يمكن أن يتولاها الوكيلُ باستقلاليةٍ عالية. الفئةُ الثانية: مهامٌّ تتطلب إشرافًا، كتحليلٍ إحصائيٍّ تختار أنت طريقتَه ويتولى الوكيلُ تنفيذَها مع مراجعتك للنتائج. الفئةُ الثالثة: مهامٌّ لا تُؤتمت، كتأويل دلالة نتيجةٍ بيولوجيةٍ أو اتخاذ قرارٍ بحثيٍّ استراتيجي. النجاحُ في نشر الوكلاء يبدأ بوضع كل مهمةٍ في فئتها الصحيحة.

حساب العائد على الاستثمار

ليست كلُّ أتمتةٍ مجديةً اقتصاديًّا. فبناءُ وكيلٍ لمهمةٍ تؤديها مرةً واحدةً في السنة قد يستهلك وقتًا أكثرَ مما يوفِّر. القاعدةُ العملية: أتمت المهامَّ التي تتكرر كثيرًا وتستغرق وقتًا معقولًا في كل مرة. مهمةٌ تؤديها يوميًّا وتأخذ عشر دقائق تستحق الأتمتة أكثرَ من مهمةٍ تؤديها شهريًّا وتأخذ ساعة. وكلما نضجت أدواتُك وقوالبُك، انخفضت تكلفةُ بناء وكيلٍ جديد، فيتسع نطاقُ ما يستحق الأتمتة. ابدأ بالمهام عالية التكرار، وراكم خبرتَك وقوالبَك تدريجيًّا.

الفصل ٣

تجهيز بيئة العمل الوكيلية

﴿ الَّذِي أَحْسَنَ كُلَّ شَيْءٍ خَلَقَهُ ﴾

السجدة: ٧

قبل أن نطلق الوكلاء على مهامنا، نحتاج إلى بيئةٍ مهيأة. وتجهيزُ هذه البيئة في سياق المعلوماتية الحيوية له اعتباراتٌ خاصة، أهمُّها التوازنُ بين القدرة والأمان، خصوصًا حين نتعامل مع بياناتٍ حساسة. سنستعرض في هذا الفصل المكوناتِ الأساسيةَ لبيئة عملٍ وكيليةٍ فعّالةٍ وآمنة.

اختيار النموذج المناسب للمهام العلمية

تختلف النماذجُ في قدراتها وتكلفتها وملاءمتها للعمل العلمي. للمهام التي تتطلب استدلالًا معقدًا — كتأويل نتيجةٍ أو تصميم تحليل — تحتاج نموذجًا قويًّا. أما للمهام الروتينية المتكررة — كتنسيق ملفٍ أو استخراج حقلٍ من نص — فقد يكفي نموذجٌ أصغر وأسرع وأرخص. ومن الحكمة بناءُ نظامٍ يوجِّه كلَّ مهمةٍ إلى النموذج الأنسب لها، فتوازن بين الجودة والتكلفة. وانتبه إلى أن النماذجَ العامةَ قد لا تعرف أحدثَ الأدوات

والمصطلحات المتخصصة، لذا قد تحتاج إلى تزويدها بمراجعَ محدَّثةٍ عبر تقنية الاسترجاع.

الأدوات الأساسية للمعلوماتي الحيوي

يحتاج وكيلُ المعلوماتية الحيوية إلى أدواتٍ تربطه بعالم البيانات البيولوجية. من أهمها: أداةٌ لتنفيذ الشيفرة (بايثون و R) في بيئةٍ معزولة، وأداةٌ للوصول إلى واجهات برمجة قواعد البيانات الحيوية مثل خدمات NCBI، وأداةٌ للبحث في الأدبيات العلمية، وأداةٌ لقراءة الملفات بصيغها البيولوجية المتنوعة (FASTA و FASTQ و VCF و BAM وغيرها). كلُّ أداةٍ يجب أن تُصمَّم بوصفٍ دقيقٍ يفهمه الوكيلُ، وبحدودٍ واضحةٍ تمنعه من تجاوز صلاحياته. تذكَّر أن جودةَ أدواتك تحدد سقفَ ما يستطيع وكيلُك إنجازه.

العزل والأمان: مبدأ أقل الصلاحيات

في بيئةٍ علميةٍ قد تتعامل مع بياناتٍ ثمينةٍ أو حساسة، يصبح العزلُ ضرورةً لا رفاهية. شغِّل شيفرةَ الوكيل في بيئةٍ معزولة (حاوية أو بيئةٍ افتراضية) لا تستطيع الوصولَ إلا إلى ما تحتاجه فعلًا. طبِّق مبدأ أقل الصلاحيات: لا تمنح الوكيلَ صلاحيةَ حذف الملفات إن كان يحتاج قراءتَها فقط، ولا وصولًا إلى الإنترنت إن كانت مهمتُه محليةً بحتة. واجعل العملياتِ التي لا رجعةَ فيها — كحذف بياناتٍ أو الكتابةِ فوق ملفاتٍ أصلية — تتطلب موافقتَك الصريحة. هذه الحواجزُ ليست عقباتٍ أمام الإنتاجية، بل ضماناتٌ تحميك من خطأٍ مكلِّف.

إدارة المفاتيح والاعتمادات

تتطلب كثيرٌ من قواعد البيانات والخدمات مفاتيحَ وصول. لا تضع هذه المفاتيح أبدًا في الشيفرة مباشرةً ولا تشاركها مع الوكيل في نصّ المحادثة، بل احفظها في متغيرات بيئةٍ آمنةٍ أو خزائنِ أسرارٍ مخصصة. وانتبه بوجهٍ خاص حين يتعامل الوكيلُ مع بياناتٍ صحيةٍ خاضعةٍ لقوانينَ كقانون HIPAA الأمريكي، إذ قد يكون إرسالُ بياناتِ مرضى إلى خدمةٍ خارجيةٍ مخالفةً قانونية. في هذه الحالات، فضِّل النماذجَ التي تعمل في بيئتك المحلية أو في بنيةٍ تحتيةٍ متوافقةٍ مع متطلبات الخصوصية. سنفصِّل هذا الجانب في فصل الأخلاقيات.

الفصل ٤

وكلاء مراجعة الأدبيات وتركيب المعرفة

﴿ فَاسْأَلُوا أَهْلَ الذِّكْرِ إِن كُنتُمْ لَا تَعْلَمُونَ ﴾

النحل: ٤٣

تبدأ كلُّ رحلةٍ بحثيةٍ بفهم ما سبقها. لكنَّ عددَ المقالات العلمية في علوم الحياة ينمو بمعدلٍ يفوق طاقةَ أيِّ باحثٍ على المتابعة: مئاتُ الآلاف من المقالات سنويًّا في قواعدَ كـ PubMed. هنا يصبح وكيلُ مراجعة الأدبيات أداةً لا غنى عنها، يساعدك على المسح السريع، وتحديد الأهم، وتركيب صورةٍ أولية. لكنه أداةٌ يجب استخدامُها بحذرٍ بالغ، كما سنرى.

البحث الذكي والتصفية

الوكيلُ المتقن لا يكتفي بتمرير سؤالك إلى محرك بحث، بل يصوغ استعلاماتٍ متعددةً بزواياَ مختلفة، ويبحث في قواعدَ متخصصةٍ كـ PubMed عبر واجهاتها البرمجية، ثم يصفّي النتائج حسب الصلة والحداثة والأهمية، ويزيل المكرر، ويصنّف ما تبقى. تخيَّل أن تطلب: «اجمع أحدثَ المقالات عن دور جينٍ معينٍ في مرضٍ معين خلال العامين

الأخيرين، وصنِّفها حسب نوع الدراسة». الوكيلُ ينفِّذ ذلك في دقائقَ معدودة، فيوفِّر عليك ساعاتٍ من البحث اليدوي، ويمنحك نقطةَ انطلاقٍ منظَّمة.

التلخيص وتركيب المعرفة

بعد الجمع يأتي التركيب. يستطيع الوكيلُ أن يقرأ ملخصاتِ عشرات المقالات ويستخرج منها الأنماطَ المشتركة، والنتائجَ المتعارضة، والفجواتِ البحثية. ويمكنه أن يبني جدولًا يقارن منهجياتِ الدراسات ونتائجها، أو أن يرسم خريطةً مفاهيميةً للحقل. هذا «التركيب» هو حيث تكمن القيمةُ الحقيقية، إذ يحوِّل كومةً من المقالات المتفرقة إلى فهمٍ منظَّمٍ قابلٍ للبناء عليه. لكن انتبه: التلخيصُ الذي يقدمه الوكيلُ هو نقطةُ بدايةٍ لقراءتك أنت، لا بديلٌ عنها.

خطر الهلوسة في السياق العلمي

أخطرُ ما يهدد وكيلَ الأدبيات هو «الهلوسة»: أن يختلق مرجعًا لا وجود له، أو ينسب نتيجةً إلى دراسةٍ لم تقلها، أو يخلط بين نتائج دراساتٍ مختلفة. وهذا خطرٌ كارثيٌّ في العلم، فاقتباسُ مرجعٍ وهميٍّ قد يدمِّر مصداقيةَ بحثٍ كامل. الوقايةُ تقوم على مبدأٍ صارم: لا تثق بأيِّ ادعاءٍ علميٍّ يقدمه الوكيلُ دون أن تتحقق من مصدره الأصلي بنفسك. استخدم تقنياتِ الاسترجاع التي تُلزم الوكيلَ بأن يبني كلامَه على نصوصٍ حقيقيةٍ مسترجَعة، واطلب منه دائمًا أن يستشهد بالمصدر الدقيق لكل ادعاء، ثم افتح المصدرَ وتأكد. القاعدةُ المقدسة: تحقَّق ثم تحقَّق.

الفصل ٥

استرجاع البيانات والاستعلام عن قواعد البيانات

﴿ وَعِندَهُ مَفَاتِحُ الْغَيْبِ لَا يَعْلَمُهَا إِلَّا هُوَ ﴾

الأنعام: ٥٩

تتوزع البياناتُ البيولوجية على شبكةٍ هائلةٍ من قواعد البيانات العامة، لكلٍّ منها واجهتُها وصيغتُها وطريقةُ استعلامها. هذا التنوعُ نعمةٌ ونقمة: نعمةٌ لأن المعرفةَ متاحةٌ للجميع، ونقمةٌ لأن استرجاع البيانات يتطلب إتقانَ أدواتٍ كثيرة. الوكيلُ المتقنُ يخفّف هذا العبء بأن يصبح وسيطًا ذكيًّا بينك وبين هذه القواعد.

قواعد البيانات الأساسية

يحتاج كلُّ معلوماتيٍّ حيويٍّ إلى الإلمام بقواعدَ أساسية. فمركزُ NCBI يضم كنوزًا كـ GenBank لتسلسلات الحمض النووي، و PubMed للأدبيات، و GEO و SRA لبيانات التعبير الجيني والتسلسل. وقاعدةُ Ensembl تقدم تعليقًا جينوميًّا غنيًّا. و UniProt مرجعٌ للبروتينات. و PDB لبنى البروتينات ثلاثية الأبعاد. ولكلٍّ من هذه

القواعد واجهةٌ برمجيةٌ تتيح الوصولَ الآليَّ إلى محتواها. والوكيلُ، حين نزوّده بأدواتٍ تتحدث مع هذه الواجهات، يستطيع أن يجلب لك ما تحتاجه بأمرٍ بلغةٍ طبيعية بدلًا من تذكُّر تفاصيل كل واجهة.

من اللغة الطبيعية إلى الاستعلام المنظَّم

أقوى ما يقدمه الوكيلُ هنا هو الترجمةُ من نيَّتك إلى استعلامٍ تقني. أنت تقول: «أريد كلَّ عيِّنات سرطان الكلى في GEO التي استُخدم فيها تسلسلُ الحمض النووي الريبي، ونُشرت بعد عامٍ معين»، فيحوِّل الوكيلُ هذا الطلبَ إلى استعلامٍ منظَّمٍ بالحقول والمرشِّحات الصحيحة، وينفِّذه، ويعيد لك النتائجَ مرتبة. هذا التحويلُ من القصد إلى التنفيذ يلغي حاجزًا تقنيًّا طالما أعاق الباحثين غيرَ المتخصصين في البرمجة، ويجعل ثرواتِ البيانات في متناول كل باحث.

التحقق من سلامة البيانات المسترجَعة

البياناتُ المسترجَعة ليست دائمًا نظيفةً أو كاملة. قد تكون هناك عيِّناتٌ ناقصةُ البيانات الوصفية، أو معرِّفاتٌ متضاربة، أو صيغٌ غيرُ متوقعة. الوكيلُ الجيد لا يكتفي بالجلب، بل يفحص ما جلبه: يتحقق من اكتمال الحقول، ويبلِّغ عن أيِّ شذوذ، ويصف لك حالةَ البيانات قبل أن تبني عليها. ومن الحكمة أن تطلب من الوكيل دائمًا تقريرًا موجزًا عن البيانات المسترجَعة: كم عيِّنة، وما مدى اكتمالها، وهل ثمة ما يدعو للانتباه. فالتحليلُ الذي يُبنى على بياناتٍ معيبةٍ ينتج نتائجَ معيبةً مهما بلغت دقةُ المعالجة.

الفصل ٦

وكلاء تحليل التسلسلات

﴿ خَلَقَ الْإِنسَانَ مِنْ عَلَقٍ ﴾

العلق: ٢

تحليلُ التسلسلات هو قلبُ المعلوماتية الحيوية النابض. فمن مقارنة تسلسلٍ جينيٍّ بقاعدة بياناتٍ ضخمة، إلى محاذاة عدة تسلسلاتٍ لكشف المناطق المحفوظة، إلى التنبؤ ببنية بروتين، تتكرر هذه المهامُّ يوميًّا في مختبرات العالم. وكثيرٌ منها يتضمن تشغيلَ أدواتٍ متخصصةٍ بمعاملاتٍ دقيقة وتفسيرَ مخرجاتها، وهو نمطٌ مثاليٌّ لمساعدة الوكلاء.

أتمتة البحث عن التشابه

من أشهر مهام تحليل التسلسلات البحثُ عن التشابه باستخدام أدواتٍ كـ BLAST، التي تقارن تسلسلَك بملايين التسلسلات المعروفة لتجد أقربَها. هذه المهمةُ تتضمن خطواتٍ متكررة: تجهيز التسلسل، واختيار قاعدة البيانات المناسبة، وضبط معاملات الحساسية، وتشغيل البحث، ثم تفسير النتائج. الوكيلُ يستطيع أن يتولى هذه السلسلةَ

كاملة: تأخذ تسلسلًا وتقول «ابحث عن أقرب التطابقات واشرح لي ماذا تعني»، فيجهِّز ويشغِّل ويفسِّر، ويبرز لك التطابقاتِ ذاتَ الدلالة مع قيم الثقة الإحصائية.

المحاذاة وتحليل المناطق المحفوظة

محاذاةُ التسلسلات المتعددة تكشف المناطقَ التي حُفظت عبر التطور، وهي غالبًا مناطقُ ذاتُ وظيفةٍ حيوية. تشغيلُ أدوات المحاذاة وتفسيرُ نتائجها وتحويلُها إلى رسومٍ توضيحية مهامُّ يتقنها الوكيل. يمكنك أن تطلب منه محاذاةَ مجموعةٍ من التسلسلات، ثم تمييزَ المواقع المحفوظة، ثم توليدَ رسمٍ يبرزها بصريًّا. وهكذا تنتقل من بياناتٍ خام إلى رؤيةٍ بيولوجيةٍ في خطواتٍ قليلة، مع بقاء التأويل النهائي — لماذا حُفظت هذه المنطقة وما دلالتها الوظيفية — بين يديك أنت بوصفك الباحث.

حدود الوكيل في التحليل المتخصص

رغم قدرة الوكلاء، ثمة حدودٌ يجب إدراكها. فالتحليلُ المتخصص قد يتطلب خبرةً عميقةً في اختيار الأداة الصحيحة والمعاملات المناسبة لنوع بياناتٍ بعينه. قد يختار الوكيلُ معاملاتٍ افتراضيةً لا تلائم حالتك، أو يطبِّق أداةً صُممت لسياقٍ مختلف. لذلك يجب أن تعرف أنت ما الذي يفعله الوكيلُ ولماذا، لا أن تسلِّمه القيادَ أعمى. استخدم الوكيلَ لتسريع التنفيذ، لكن احتفظ بالقرار المنهجي لنفسك، وراجع المعاملاتِ التي اختارها قبل أن تعتمد النتائج.

الفصل ٧

تنسيق خطوط الأنابيب وأتمتة سير العمل

﴿ إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ ﴾

القمر: ٤٩

نادرًا ما يكتمل تحليلٌ في المعلوماتية الحيوية بأداةٍ واحدة. الواقعُ أن معظم التحليلات سلسلةٌ من الخطوات، كلٌّ منها يأخذ مخرجَ سابقتها مدخلًا له: من ضبط الجودة، إلى المحاذاة، إلى العدِّ، إلى التحليل الإحصائي. هذه السلسلةُ تُسمَّى «خط الأنابيب» (Pipeline)، وإدارتُها بكفاءةٍ وقابليةِ إعادة إنتاجٍ هي من أصعب التحديات اليومية وأكثرها مكافأةً عند إتقانها.

أنظمة إدارة سير العمل

طوَّر مجتمعُ المعلوماتية الحيوية أنظمةً متخصصةً لإدارة خطوط الأنابيب، أشهرُها Nextflow و Snakemake. هذه الأنظمةُ تتيح وصفَ التحليل بوصفه سلسلةً من القواعد المترابطة، فتدير التنفيذَ تلقائيًّا، وتوازي الخطواتِ المستقلة، وتعيد التشغيلَ من حيث توقف عند الفشل، وتضمن إعادةَ الإنتاج. لكنَّ كتابة هذه السكربتات وصيانتها

تتطلب خبرة. وهنا يبرز الوكيل: يستطيع أن يساعدك في كتابة قاعدةٍ جديدة، أو تصحيح خطأٍ في سير العمل، أو تحويل تحليلٍ يدويٍّ متفرقٍ إلى خط أنابيبَ منظَّمٍ قابلٍ للتكرار.

الوكيل كمنسِّق لسير العمل

أبعدُ من مجرد كتابة السكربتات، يمكن للوكيل أن يكون منسِّقًا حيًّا لسير العمل. يراقب تنفيذَ خط الأنابيب، ويكتشف حين تفشل خطوة، ويقرأ رسالةَ الخطأ، ويقترح إصلاحًا، بل وقد ينفِّذ الإصلاحَ ويعيد التشغيل بعد موافقتك. هذا النمطُ من «الإشراف الذكي» يحرِّر الباحثَ من مراقبة التحليلات الطويلة ساعةً بساعة، ويتيح له أن يتدخل فقط حين يلزم قرارٌ بشري. لكن احرص دائمًا على أن تبقى لك الكلمةُ الأخيرة في أيِّ تغييرٍ يؤثر على منهجية التحليل أو نتائجه.

قابلية إعادة الإنتاج كمبدأ أول

إنَّ أعظمَ ما يمكن أن يقدمه الوكيلُ لسير العمل ليس السرعة، بل قابليةُ إعادة الإنتاج. فالعلمُ الذي لا يمكن تكرارُه ليس علمًا. اطلب من وكيلك دائمًا أن يوثِّق كلَّ خطوة: إصداراتِ الأدوات المستخدمة، والمعاملاتِ المحددة، ومصدرَ كل بيان. واجعله يبني خطوطَ أنابيبَ مكتفيةً ذاتيًّا، تستطيع أنت أو غيرُك تشغيلَها لاحقًا فتعطي النتائجَ نفسها. الوكيلُ المنضبط الذي يجعل التوثيقَ وقابليةَ إعادة الإنتاج عادةً متأصلةً يحمي بحثَك من أكبر آفات العلم الحاسوبي: النتائجُ التي لا يستطيع أحدٌ — ولا حتى صاحبُها — أن يكررها.

الفصل ٨

معالجة البيانات والتحليل الإحصائي

﴿ أَحْصَىٰ كُلَّ شَيْءٍ عَدَدًا ﴾

الجن: ٢٨

يقول كثيرٌ من المعلوماتيين الحيويين إن ثمانين بالمئة من وقتهم يضيع في تنظيف البيانات وتجهيزها، وعشرين بالمئة فقط في التحليل الفعلي. هذه «المعالجة» (Data Wrangling) — من توحيد الصيغ، إلى التعامل مع القيم المفقودة، إلى دمج الجداول — مملةٌ ومتكررةٌ ومعرَّضةٌ للخطأ، وهي بالضبط ما يبرع الوكلاءُ في تخفيف عبئه.

تنظيف البيانات وتجهيزها

يستطيع الوكيلُ أن يفحص مجموعةَ بياناتٍ ويصف لك حالتها: كم صفًّا وعمودًا، وأين القيمُ المفقودة، وما الأنماطُ الشاذة. ثم يكتب الشيفرةَ — بلغة بايثون أو R — لتنظيفها: توحيدُ أسماء العيِّنات، وتصحيحُ الأنواع، ومعالجةُ القيم المفقودة بالطريقة التي تختارها، ودمجُ مصادرَ متعددة في جدولٍ واحدٍ متسق. وأهمُّ ما في الأمر أن الوكيلَ يجعل هذه

المعالجةَ مكتوبةً في شيفرةٍ قابلةٍ للمراجعة وإعادة التشغيل، بدلًا من تعديلاتٍ يدويةٍ خفيةٍ في جدولٍ إلكترونيٍّ يستحيل تتبعُها أو تكرارُها.

التحليل الإحصائي تحت الإشراف

في التحليل الإحصائي يتعمق دورُ الإشراف البشري. يستطيع الوكيلُ أن ينفِّذ تحليلاتٍ معقدة — كتحليل التعبير الجيني التفاضلي، أو التحليل العنقودي، أو تحليل المكوِّنات الرئيسة — ويولِّد الشيفرةَ والمخرجات. لكنَّ اختيارَ الطريقة الإحصائية الصحيحة، والتأكدَ من تحقُّق فرضياتها، وتأويلَ دلالة النتائج، كلُّها قراراتٌ علميةٌ يجب أن تتخذها أنت. خطرٌ شائع أن يطبِّق الوكيلُ اختبارًا إحصائيًّا لا تتحقق شروطُه على بياناتك، فيعطي نتيجةً تبدو سليمةً لكنها مضلِّلة. لذا: دع الوكيلَ ينفِّذ الحساب، لكن تولَّ أنت اختيارَ المنهج والحكمَ على صحته.

تحليل البيانات أحادية الخلية كمثال

خذ تحليلَ بيانات التسلسل أحادي الخلية مثالًا على التعقيد الحديث. يتضمن خطواتٍ كثيرة: ضبطُ الجودة، والتطبيعُ، وتقليلُ الأبعاد، والتجميعُ العنقودي، وتحديدُ أنواع الخلايا. وكلُّ خطوةٍ فيها خياراتٌ ومعاملاتٌ تؤثر على النتيجة. يستطيع الوكيلُ أن يبني خطَّ التحليل بأكمله باستخدام مكتباتٍ معيارية، ويولِّد الرسومَ في كل مرحلة، ويشرح ما يحدث. لكنه قد يخطئ في تحديد أنواع الخلايا أو في اختيار عتبةٍ ما. فاستفد من سرعته في التنفيذ، وراجع كلَّ قرارٍ بيولوجيٍّ بعينك الخبيرة، فالتحليلُ المتقن شراكةٌ بين سرعة الآلة وبصيرة الباحث.

الفصل ٩

توليد الشيفرة والتصحيح وقابلية إعادة الإنتاج

﴿ صُنْعَ اللَّهِ الَّذِي أَتْقَنَ كُلَّ شَيْءٍ ﴾

النمل: ٨٨

البرمجةُ عملةُ المعلوماتية الحيوية اليومية. فكلُّ تحليلٍ يتطلب شيفرةً: سكربتاتٌ لمعالجة البيانات، ودوالُّ للحسابات، وأوامرُ لتشغيل الأدوات. وقد أحدثت وكلاءُ البرمجة ثورةً هنا، إذ صارت قادرةً لا على اقتراح أسطرٍ من الشيفرة فحسب، بل على كتابة وحداتٍ كاملة، وتصحيح أخطاءٍ متشعبة، وإعادة هيكلة شيفرةٍ قائمة. هذا الفصلُ عن كيفية الاستفادة من هذه القدرة بأمانٍ وإتقان.

من الوصف إلى الشيفرة العاملة

تستطيع أن تصف للوكيل ما تريد بلغةٍ طبيعية — «اكتب دالةً تقرأ ملفَّ VCF، وتصفّي المتغيّرات حسب الجودة، وتعيد جدولًا بالمتغيّرات الباقية» — فيكتب الشيفرةَ كاملةً، ويشرحها، ويقترح اختباراتٍ لها. والأقوى من ذلك أن وكلاءَ البرمجة الحديثة تعمل في دورة: تكتب الشيفرة، وتشغّلها، وتقرأ الخطأَ إن وُجد، وتصحّح، وتعيد، حتى

تنجح. هذه القدرةُ على إغلاق حلقة «اكتب-شغِّل-صحِّح» ذاتيًّا هي ما يميز الوكيلَ عن مجرد مولِّد شيفرة، وتجعله شريكًا برمجيًّا حقيقيًّا.

خطر الشيفرة التي تعمل لكنها خاطئة

أخطرُ ما في الشيفرة المولَّدة ليس الخطأَ الذي يوقف التشغيل، بل الخطأُ الصامت: شيفرةٌ تعمل دون أن تتعطل، لكنها تحسب الشيءَ الخطأ. قد يخلط الوكيلُ بين تعريفين لمقياسٍ ما، أو يطبِّق تحويلًا في غير موضعه، فتخرج أرقامٌ تبدو معقولةً لكنها مغلوطة. وهذا في العلم كارثة، لأن نتائجَك ستُبنى على حسابٍ خاطئ دون أن تدري. الحمايةُ تكون باختبار الشيفرة على حالاتٍ تعرف نتائجَها سلفًا، وبمراجعة المنطق لا مجرد التأكد من أنها «تعمل». لا تثق بأن الشيفرة صحيحةٌ لمجرد أنها أنتجت رقمًا؛ تأكد أنها أنتجت الرقمَ الصحيح.

البيئات المعزولة والإصدارات

لضمان أن تعمل شيفرتُك اليوم كما تعمل بعد عام، احرص على عزل بيئاتها وتثبيت إصداراتها. اطلب من الوكيل أن يحدد بدقة إصداراتِ كل مكتبةٍ تستخدمها، وأن يبني بيئةً معزولةً يمكن إعادةُ بنائها لاحقًا. وضع شيفرتَك تحت نظام تحكُّمٍ بالإصدارات، واجعل الوكيلَ يساعدك في كتابة رسائل التغيير الواضحة. هذه العاداتُ تبدو شكليةً لكنها جوهرُ العلم الحاسوبي القابل لإعادة الإنتاج. الوكيلُ الذي يعينك على الانضباط هنا يقدِّم لبحثك خدمةً أعظمَ من أيِّ تسريعٍ آني.

الفصل ١٠

وكلاء التصوير وإعداد التقارير

﴿ قُلِ انظُرُوا مَاذَا فِي السَّمَاوَاتِ وَالْأَرْضِ ﴾

يونس: ١٠١

النتيجةُ التي لا تُعرَض جيدًا نتيجةٌ مهدورة. ففي المعلوماتية الحيوية، حيث الأرقامُ كثيرةٌ والعلاقاتُ معقدة، يكون التصويرُ الجيد هو الجسرَ بين التحليل والفهم، وبين الباحث وجمهوره. وإعدادُ الرسوم والتقارير، رغم أهميته، يستهلك وقتًا طويلًا في الضبط والتنسيق. وهنا يقدم الوكلاءُ عونًا كبيرًا.

توليد الرسوم العلمية

يستطيع الوكيلُ أن يحوِّل وصفًا بلغةٍ طبيعيةٍ إلى رسمٍ علميٍّ متقن. تقول: «ارسم خريطةً حراريةً لتعبير هذه الجينات عبر العيِّنات، مع تجميعٍ هرميٍّ للصفوف والأعمدة»، فيكتب الشيفرةَ، ويولِّد الرسم، ويضبط الألوانَ والمقاييسَ والعناوين. ويمكنك أن تطلب تعديلاتٍ متتابعةً بلغةٍ طبيعية: «كبِّر الخط»، «غيِّر تدرُّجَ الألوان»، «أضف فاصلًا بين

المجموعات». هذا الحوارُ التفاعلي يلغي دوراتِ التجربة والخطأ المملة في ضبط الرسوم، ويتيح لك الوصولَ إلى الشكل المطلوب بسرعة.

التقارير القابلة لإعادة الإنتاج

أرقى ما في إعداد التقارير أن تكون قابلةً لإعادة الإنتاج: مستنداتٌ تجمع بين النص والشيفرة والمخرجات، بحيث يُعاد توليدُ التقرير كاملًا حين تتغير البيانات. أدواتٌ كـ R Markdown و Jupyter و Quarto تتيح ذلك، والوكيلُ يبرع في بنائها. يمكنك أن تطلب تقريرًا يحلّلِ بياناتك ويولِّد الرسومَ والجداولَ والتفسيرَ في مستندٍ واحدٍ متكامل. وحين تصلك بياناتٌ جديدة، يُعاد توليدُ التقرير بضغطة زر. هذا النمطُ يجمع بين الكفاءة وقابلية إعادة الإنتاج، وهو من أنفع ما يقدمه الوكيل للباحث.

حدود التفسير الآلي للنتائج

قد يولِّد الوكيلُ، إلى جانب الرسوم، تفسيراتٍ نصيةً للنتائج. وهنا تجب الحيطة. فالوكيلُ قد يصف ما يراه في الرسم وصفًا سطحيًّا صحيحًا، لكنه قد يقفز إلى استنتاجاتٍ بيولوجيةٍ لا يسندها دليل، أو يضفي على ارتباطٍ إحصائيٍّ معنًى سببيًّا لا يحق له. التفسيرُ العلمي العميق — ربطُ النتيجة بالسياق البيولوجي، وتقديرُ دلالتها، وتمييزُ الارتباط من السببية — يبقى مسؤوليةَ الباحث. استخدم وصفَ الوكيل للرسوم نقطةَ انطلاقٍ، لكن اكتب التفسيرَ العلمي بنفسك، فهو جوهرُ إسهامك الذي لا يُفوَّض.

الفصل ١١

أنظمة الوكلاء المتعددة للتحليلات المعقدة

﴿ وَأَمْرُهُمْ شُورَىٰ بَيْنَهُمْ ﴾

الشورى: ٣٨

حين يتعقد التحليلُ ويتشعب، قد لا يكفي وكيلٌ واحدٌ يحاول فعلَ كل شيء. تخيَّل مشروعًا يتطلب مراجعةَ أدبيات، واسترجاعَ بيانات، وتحليلًا إحصائيًّا، وتصويرًا، وكتابةَ تقرير. توزيعُ هذه الأدوار على وكلاءَ متخصصين، يتشاورون ويتكاملون، قد يكون أنجعَ من تحميل وكيلٍ واحدٍ كلَّ هذا العبء. هذا هو منطقُ أنظمة الوكلاء المتعددة في سياق المعلوماتية الحيوية.

فرق الوكلاء المتخصصة

يمكن تصميمُ فريقٍ من الوكلاء يحاكي فريقَ بحثٍ بشري. وكيلٌ متخصصٌ في البحث يجمع الأدبيات والبيانات. ووكيلٌ متخصصٌ في التحليل يكتب الشيفرةَ وينفِّذ الحسابات. ووكيلٌ متخصصٌ في المراجعة يفحص مخرجاتِ الآخرين بحثًا عن الأخطاء. ووكيلٌ منسِّقٌ يوزِّع المهامَّ ويجمع النتائج. هذا التخصصُ يتيح لكل وكيلٍ أن يحمل في

سياقه ما يخص مهمته فقط، فلا يزدحم بكل تفاصيل المشروع، ويؤدي دورَه بإتقانٍ أعلى. والأهمُّ أن وكيلَ المراجعة يضيف طبقةَ تحقُّقٍ تكشف أخطاءَ قد تفلت من وكيلٍ منفرد.

التحقق المتبادل كطبقة أمان

من أثمن فوائد تعدد الوكلاء في العلم آليةُ التحقق المتبادل. فحين يراجع وكيلٌ عملَ آخر، تنشأ رقابةٌ تكشف الأخطاءَ مبكرًا. يمكن مثلًا أن يكتب وكيلٌ تحليلًا إحصائيًّا، ثم يفحصه وكيلٌ آخرُ متسائلًا: هل الطريقةُ مناسبة؟ هل الفرضياتُ متحققة؟ هل التفسيرُ مدعوم؟ هذه المساءلةُ الآلية، رغم أنها لا تغني عن مراجعة الباحث، تضيف خطَّ دفاعٍ إضافيًّا ضد الأخطاء الصامتة التي تهدد دقةَ العلم الحاسوبي.

متى تتجنب التعقيد؟

رغم جاذبية الفكرة، التحذيرُ نفسُه يتكرر: لا تلجأ إلى تعدد الوكلاء إلا حين تعجز بنيةٌ أبسطُ عن حل المشكلة. فأنظمةُ الوكلاء المتعددة أعقدُ في البناء والصيانة، وأغلى في التكلفة، وأصعبُ في تتبع الأخطاء. كثيرٌ من المهام التي تبدو محتاجةً لفريقٍ كامل، يحلها وكيلٌ واحدٌ مصمَّمٌ جيدًا بأدواتٍ مناسبة. القاعدةُ في المعلوماتية الحيوية كما في غيرها: ابدأ بسيطًا، وأضف التعقيدَ فقط حين يثبت أن البساطةَ لا تكفي. التعقيدُ ثمنٌ يُدفع عند الضرورة، لا غايةٌ تُطلب لذاتها.

الفصل ١٢

التحقق والأمان والصرامة العلمية

﴿ إِن جَاءَكُمْ فَاسِقٌ بِنَبَإٍ فَتَبَيَّنُوا ﴾

الحجرات: ٦

هذا الفصلُ هو ضميرُ الكتاب. فكلُّ ما سبق من قدراتٍ يصبح خطرًا إن لم يُضبط بصرامةٍ علمية. الآيةُ الكريمة تضع القاعدةَ الذهبية: إذا جاءك خبرٌ من مصدرٍ قد لا يكون موثوقًا، فتبيَّن قبل أن تبني عليه. والوكيلُ، مهما بلغت قدرتُه، مصدرٌ يجب التبيُّنُ من مخرجاته، لا سلطةٌ تُقبل بلا فحص.

الهلوسة في العلوم الحيوية

تأخذ الهلوسةُ في المعلوماتية الحيوية أشكالًا خبيثة. قد يخترع الوكيلُ اسمَ جينٍ أو بروتينٍ لا وجود له، أو ينسب وظيفةً خاطئةً لجينٍ حقيقي، أو يستشهد بمقالةٍ وهمية، أو يدَّعي وجودَ ارتباطٍ بين مرضٍ وعاملٍ بلا سند. وخطورتُها أنها تأتي بثقةٍ تامةٍ وصياغةٍ علميةٍ مقنعة. لا يوجد حلٌّ سحريٌّ يلغي الهلوسةَ تمامًا، لكن تخفيفَها ممكنٌ بإلزام الوكيل

بالاعتماد على مصادرَ مسترجَعةٍ حقيقية، وبطلب الاستشهاد الدقيق، وقبل كل شيء بالتحقق البشري المستقل من كل ادعاءٍ جوهري.

التحقق المنهجي من المخرجات

التحققُ ليس شعورًا بل منهجٌ منظَّم. لكلِّ نوعٍ من المخرجات طريقةُ تحقُّقٍ تناسبه. الشيفرةُ تُختبَر على حالاتٍ معروفة النتائج. النتائجُ الإحصائية تُراجَع فرضياتُها وتُعاد بطريقةٍ مستقلة عند الشك. الادعاءاتُ العلمية تُرَدُّ إلى مصادرها الأصلية. الأرقامُ تُفحَص هل هي معقولةٌ بيولوجيًّا. اجعل التحققَ خطوةً ثابتةً في سير عملك، لا فكرةً لاحقة. وكلما زادت أهميةُ النتيجة وعواقبُها، زادت صرامةُ التحقق المطلوب. هذه الصرامةُ ليست عائقًا للسرعة، بل هي ما يفصل العلمَ عن مجرد توليدٍ مقنعٍ للنصوص.

تجنب التحيز في التحليل

ثمة خطرٌ أدقُّ من الخطأ الصريح، وهو التحيز. قد يميل الباحثُ — بوعيٍ أو بغير وعي — إلى قبول مخرجات الوكيل التي تؤكد فرضيتَه، وإهمال ما يخالفها. والوكيلُ نفسُه قد يحمل تحيزاتٍ من بيانات تدريبه. الوقايةُ تكون بأن تحدد معاييرَ النجاح قبل أن ترى النتائج، وأن تختبر فرضياتِك بحثًا عن دحضها لا عن تأكيدها، وأن تطلب من الوكيل أحيانًا أن يقدم الحجةَ المضادة. الصرامةُ العلمية تتطلب أن نكون أقسى على الأفكار التي نحبها، لا أن نبحث عمن يصادق عليها.

الفصل ١٣

دراسة حالة: تحليل متكامل بالوكلاء

﴿ فَامْشُوا فِي مَنَاكِبِهَا ﴾

الملك: ١٥

بعد كل النظرية والمبادئ، لنجمعها في سيناريو عمليٍّ متكامل. سنتتبع باحثًا افتراضيًّا يريد تحليلَ بيانات تعبيرٍ جينيٍّ لمقارنة عيِّناتٍ مريضةٍ بأخرى سليمة، مستعينًا بالوكلاء في كل خطوة، مع إبقاء الحكم العلمي بيده. هذه الرحلةُ تجسِّد كيف تتكامل الفصولُ السابقة في عملٍ واحد.

من السؤال إلى البيانات

يبدأ الباحثُ بسؤالٍ بيولوجي: أيُّ الجينات يختلف تعبيرُها بين المرضى والأصحاء؟ يكلِّف وكيلَ الأدبيات بمسح ما نُشر عن المرض وآلياته الجزيئية، فيحصل على خلفيةٍ منظَّمةٍ خلال دقائق، يراجعها بنفسه ويتحقق من مصادرها. ثم يطلب من وكيل البيانات أن يبحث في قاعدة GEO عن مجموعة بياناتٍ مناسبة، فيسترجع الوكيلُ المرشَّحات، ويصف

حالةَ كلٍّ منها، فيختار الباحثُ المجموعةَ الأنسب بناءً على حكمه. لاحظ كيف ينفِّذ الوكيلُ العملَ الثقيل بينما يبقى القرارُ العلمي للباحث.

من البيانات إلى النتائج

بعد اختيار البيانات، يطلب الباحثُ من وكيل التحليل بناءَ خط أنابيبَ قابلٍ لإعادة الإنتاج: ضبطُ الجودة، فالتطبيع، فتحليلُ التعبير التفاضلي بطريقةٍ إحصائيةٍ يحددها الباحثُ بنفسه بناءً على طبيعة البيانات. ينفِّذ الوكيلُ، ويولِّد قائمةَ الجينات المختلفة التعبير مع قيم الدلالة. يراجع الباحثُ المعاملاتِ والفرضياتِ الإحصائية، ويتحقق من معقولية النتائج. ثم يطلب من وكيل التصوير رسومًا توضيحية: خريطةً حرارية، ورسمُ بركان، وتحليلُ مكوِّناتٍ رئيسة، فتتجسد النتائجُ في صورٍ تكشف الأنماط.

من النتائج إلى المعرفة

النتائجُ الخام ليست الغاية، بل فهمُها. يستعين الباحثُ بالوكيل في تحليل الإثراء الوظيفي لمعرفة المسارات البيولوجية التي تنتمي إليها الجينات المتغيرة، فيحصل على قائمةٍ بالمسارات المرشَّحة. لكنَّ التفسيرَ النهائي — أيُّ هذه المسارات ذو دلالةٍ حقيقيةٍ في سياق المرض، وكيف يرتبط بالفرضية الأصلية، وما الذي يستحق متابعةً تجريبية — هذا هو إسهامُ الباحث الذي لا يُفوَّض. هنا يفترق الباحثُ المتقن عمن يكتفي بمخرجات الوكيل: الأولُ يبني معرفةً، والثاني يجمع أرقامًا.

من المعرفة إلى التقرير

أخيرًا، يطلب الباحثُ من الوكيل بناءَ تقريرٍ قابلٍ لإعادة الإنتاج يجمع المنهجيةَ والشيفرةَ والنتائجَ والرسوم في مستندٍ واحد، موثِّقًا كلَّ إصدارٍ ومعامل. يكتب الباحثُ بنفسه قسمَ التفسير والمناقشة، فهو لبُّ إسهامه. والنتيجةُ عملٌ أُنجز في أيامٍ بدل أسابيع، دون التفريط في الصرامة، وقابلٌ لأن يكرره أيُّ باحثٍ آخر. هذه هي الشراكةُ المثلى: الوكيلُ يضاعف الإنتاجية، والباحثُ يضمن الجودةَ والمعنى. وهذا هو جوهرُ ما أردنا تعليمَه في هذا الكتاب.

الفصل ١٤

الأخلاقيات وخصوصية البيانات والمستقبل

﴿ وَلَا تَجَسَّسُوا ﴾

الحجرات: ١٢

تختم رحلتُنا بما يجب أن يكون في صدارة وعي كل باحث: الأخلاقياتُ والخصوصية. فالبيانات في علوم الحياة ليست أرقامًا مجردة، بل كثيرٌ منها يمسُّ خصوصيةَ بشرٍ ائتمنونا على أغلى ما يملكون: معلوماتِهم الصحية وشيفرتِهم الوراثية. ونشرُ الوكلاء على هذه البيانات يفرض علينا مسؤوليةً مضاعفة.

خصوصية البيانات الصحية

تخضع البياناتُ الصحية لقوانينَ صارمةٍ كقانون HIPAA في الولايات المتحدة، الذي يحمي معلوماتِ المرضى الخاصة. والخطرُ الجوهري أن إرسال بياناتٍ صحيةٍ خاصةٍ إلى خدمة ذكاءٍ اصطناعيٍّ خارجية قد يكون انتهاكًا قانونيًّا وأخلاقيًّا. لذلك، حين تتعامل مع بياناتٍ تحدِّد هويةَ المرضى، تجنَّب إرسالَها إلى خدماتٍ خارجية، وفضِّل نماذجَ تعمل في بيئتك المحلية أو في بنيةٍ تحتيةٍ متوافقةٍ مع متطلبات الخصوصية، أو أزِل

محدِّداتِ الهوية قبل أيِّ معالجة. الآيةُ الكريمة «ولا تجسسوا» تؤسس لمبدأٍ خالد: أن خصوصيةَ الناس حرمةٌ لا تُنتهك، ولو باسم العلم.

المسؤولية والشفافية العلمية

حين تستخدم الوكلاءَ في بحثك، تبقى المسؤوليةُ العلمية كاملةً على عاتقك أنت لا على الأداة. أنت من يضمن صحةَ النتائج، وأنت من يُساءل عنها. ومن الأمانة العلمية أن تكون شفافًا بشأن استخدامك للذكاء الاصطناعي: أن توثِّق في منهجيتك أين استخدمته وكيف تحققتَ من مخرجاته. هذه الشفافيةُ ليست اعترافًا بنقص، بل التزامٌ بالنزاهة العلمية التي تتيح للآخرين تقييمَ عملك ومراجعته. والعلمُ، في جوهره، مشروعٌ جماعيٌّ يقوم على الثقة والقابلية للتحقق.

مستقبل المعلوماتية الحيوية الوكيلية

إلى أين يتجه هذا المجال؟ نتوقع وكلاءَ أكثرَ تخصصًا في العلوم الحيوية، مدرَّبين على فهم أعمق للبيولوجيا وأدواتها. ونتوقع تكاملًا أوثق بين الوكلاء وقواعد البيانات والأدوات عبر بروتوكولاتٍ موحَّدة، فيصبح نسجُ تحليلٍ معقدٍ أيسر. وقد نرى وكلاءَ يقترحون فرضياتٍ ويصممون تجاربَ افتراضية. لكنَّ جوهرَ العلم سيبقى إنسانيًّا: السؤالُ الأصيل، والحدسُ المبدع، والحكمُ الأخلاقي، والمسؤوليةُ عن المعرفة. الوكيلُ سيكبِّر قدراتِنا، لكنه لن يحلَّ محلَّ الفضول العلمي الذي يحرِّك البحث.

وتبقى دعوتي الأخيرة: أن تكون أمتنا صانعةً لهذه الأدوات لا مستهلكةً لها فحسب. فلدينا من العقول والكفاءات ما يؤهلنا للإسهام في تشكيل هذا المستقبل، إن نحن تعلَّمنا وبنينا وأتقنّا. وما هذا الكتابُ إلا بذرةٌ نرجو أن تثمر باحثين يسخِّرون أحدثَ الأدوات لخدمة العلم والإنسان. والله الموفِّق.

الخاتمة

﴿ رَبَّنَا آتِنَا مِن لَّدُنكَ رَحْمَةً وَهَيِّئْ لَنَا مِنْ أَمْرِنَا رَشَدًا ﴾

الكهف: ١٠

وصلنا إلى ختام رحلةٍ عمليةٍ في توظيف الوكلاء الأذكياء في يوميات المعلوماتي الحيوي. انتقلنا من فهم لماذا تناسب هذه الأدواتُ طبيعةَ عملنا، إلى تجهيز البيئة، فالبحث في الأدبيات واسترجاع البيانات وتحليل التسلسلات وتنسيق خطوط الأنابيب وتوليد الشيفرة والتصوير، وصولًا إلى دراسة حالةٍ متكاملة. وإذا كان ثمة خلاصةٌ واحدة، فهي أن الوكيلَ يضاعف إنتاجيةَ الباحث حين يُستخدم في موضعه الصحيح، وبحذرٍ علميٍّ يصون دقةَ النتائج.

إن أكبر خطرٍ في نشر هذه الأدوات ليس عجزَها، بل الثقةُ العمياء بها. فالوكيلُ قد يكتب سكربتًا يعمل لكنه يحمل خطأً منطقيًّا خفيًّا، أو يلخّص مقالةً علميةً تلخيصًا مضلِّلًا، أو يقترح تفسيرًا بيولوجيًّا لا يسنده دليل. لذلك تبقى عينُ الباحث الناقدة هي خطَّ الدفاع الأخير، وتبقى قابليةُ إعادة الإنتاج والتحقق المستقل هما معيارَ العلم الحق.

وأرجو أن يخرج القارئ بقناعةٍ راسخة: أن المستقبل ليس للباحث الذي تحلُّ الآلةُ محلَّه، بل للباحث الذي يُتقن العملَ معها. فمن يجيد توجيه الوكلاء، والتحقق من مخرجاتها،

ودمجها في سير عمله، سيُنجز في يومٍ ما كان يحتاج أسبوعًا، وسيتفرغ لما يتطلب الإبداعَ والحدسَ العلميَّ الذي لا تملكه الآلة.

أسأل الله أن ينفع بهذا العمل كلَّ باحثٍ يسعى لخدمة العلم والإنسانية، وأن يجعله علمًا يُنتفع به، وأن يوفِّق أمتنا لتكون صانعةً لأدوات المعرفة لا مستهلكةً لها. وما توفيقي إلا بالله. والله أعلم.

المراجع العلمية

تفسير القرآن الكريم وعلومه

ابن كثير، إسماعيل بن عمر. تفسير القرآن العظيم. دار طيبة للنشر والتوزيع.

السعدي، عبد الرحمن بن ناصر. تيسير الكريم الرحمن في تفسير كلام المنان. مؤسسة الرسالة.

الطبري، محمد بن جرير. جامع البيان عن تأويل آي القرآن. دار هجر.

المعلوماتية الحيوية وعلم الجينوم

Pevsner, J. (2015). Bioinformatics and Functional Genomics (3rd ed.). Wiley-Blackwell.

Mount, D. W. (2004). Bioinformatics: Sequence and Genome Analysis (2nd ed.). CSHL Press.

Altschul, S. F., et al. (1990). Basic Local Alignment Search Tool (BLAST). Journal of Molecular Biology.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology.

النماذج اللغوية الكبيرة والوكلاء في العلوم

Yao, S., et al. (2023). ReAct: Synergizing Reasoning and Acting in Language Models. ICLR.

Lewis, P., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS.

Boiko, D. A., et al. (2023). Autonomous chemical research with large language models. Nature.

M. Bran, A., et al. (2024). Augmenting large language models with chemistry tools. Nature Machine Intelligence.

البروتوكولات والممارسات الهندسية والأخلاقية

Anthropic. (2024). Introducing the Model Context Protocol (MCP). Anthropic Documentation.

Anthropic. (2024). Building Effective Agents. Anthropic Engineering Blog.

Sandve, G. K., et al. (2013). Ten Simple Rules for Reproducible Computational Research. PLOS Computational Biology.

U.S. Dept. of Health & Human Services. (1996). Health Insurance Portability and Accountability Act (HIPAA).

English Section

القسم الإنجليزي

Introduction

❦ *In the name of God, the Most Gracious, the Most Merciful* ❦

Bioinformatics is living through an unprecedented transformative moment. Since the human genome sequence became available, the volume of biological data has exploded beyond the capacity of any individual researcher to absorb: genetic sequences of billions of bases, gene-expression experiments for thousands of cells, and databases that grow every day. Amid this flood, the bioinformatician spends most of their day on repetitive, ramified tasks: searching the literature, extracting data from databases, writing scripts to process numbers, debugging code, and preparing reports. This is precisely where intelligent agents come in.

This book is not a general theoretical introduction to artificial intelligence, but a practical guide aimed at those who actually work in bioinformatics. Its goal is to answer a specific question: how do I deploy intelligent agents in my daily tasks to work faster and more accurately without sacrificing scientific rigor? We move from principle to direct application: from searching the NCBI, Ensembl,

and UniProt databases, to analyzing nucleic acid sequences, to orchestrating analysis pipelines, to generating and debugging code, to visualization and report preparation.

But this book carries a warning as much as it carries a promise. Biological data is sensitive, and some of it is protected health information belonging to patients that the law safeguards. An erroneous conclusion in a science that touches human health can have grave consequences. Deploying agents in scientific research therefore requires doubled caution: verifying every output, preserving data privacy, and ensuring the reproducibility of results. The agent in bioinformatics is a powerful assistant, not a substitute for the researcher's judgment and responsibility.

I wrote this book from the position of a practitioner, a researcher in computational biology who uses these tools in daily work. I was careful to link every chapter to a real task the reader faces, not to a distant abstraction. I opened every chapter with a noble verse that reminds us that studying life in its cells and genes is, in essence, a contemplation of the wondrous creation of God, and that beneficial knowledge is both worship and responsibility.

I invite you to read this book at your screen, trying what you read on your own data and tasks. Mastery comes through practice, and the agent you build for your own task is more useful than a thousand theoretical examples. Let us begin, with the blessing of God.

Chapter 1

Why Intelligent Agents in Bioinformatics?

Bioinformatics is a field at the intersection of biology, computation, and statistics, and it is among the most time-draining fields in repetitive technical details. The typical researcher may spend hours formatting a data file so that some program will accept it, recalling the syntax of a command-line instruction, or searching for the reason behind a cryptic error message. These tasks, though necessary, do not require the scientific creativity for which the researcher was trained. Herein lies the great opportunity for intelligent agents.

The Nature of Computational Biology Work

Work in bioinformatics has characteristics that make it an ideal candidate for intelligent automation. First, the abundance of repetitive tasks that can be precisely described: downloading data, converting formats, running tools with specific parameters. Second,

the constant need to connect multiple tools and databases, since an analysis is rarely completed with a single tool. Third, the enormous, constantly changing volume of knowledge: new papers every day, tools being updated, databases growing. These three characteristics, repetition, interconnection, and the flow of knowledge, are exactly what agents excel at.

What Does an Agent Add Beyond a Simple Assistant?

The researcher may ask: isn't a language assistant I ask and that answers me enough? The difference is fundamental. A simple assistant suggests a command for you to execute yourself, whereas an agent executes it, reads its result, and builds upon it. Imagine the difference between the assistant telling you "try this command to search NCBI" and the agent actually searching, reading the results, filtering them, and summarizing them for you in a report. The agent closes the loop between idea and execution, transforming from an advisor into an executor working under your supervision.

Where Agents Fit and Where They Do Not

Not everything in research is suitable for automation. Agents are excellent at well-defined tasks: data retrieval, format conversion, generating routine code, preliminary literature search. But they are weak, and sometimes dangerous, in core scientific decisions: choosing the hypothesis, interpreting biological results, and judging the significance of a finding. The golden rule is to leave the heavy mechanical work to the agent and reserve scientific judgment for yourself. The agent saves you time and effort so you can focus your energy where your true value as a researcher lies.

Chapter 2

Mapping Daily Tasks and Where to Deploy Agents

Before we deploy any agent, we must draw a clear map of our day: which tasks do we repeat? How long do they take? Which of them drains time it does not deserve? This simple analysis is the cornerstone, because intelligent automation begins with knowing what is worth automating. Let us survey the typical tasks of a bioinformatician's day and determine where the agent adds the greatest value.

The Research Project Life Cycle

Any project in bioinformatics passes through similar stages. It begins with a literature review to understand what has been done and identify the gap. Then comes collecting and retrieving data from public databases or laboratory experiments. This is followed by the stage of processing, cleaning, and quality-controlling data, one of the most time-consuming stages. Then the primary and statistical

analysis, then visualization and interpretation of results, and finally writing the report or scientific paper. At each of these stages there are routine tasks an agent can take on.

Classifying Tasks by Automation Suitability

Tasks can be classified into three categories. The first: tasks fully safe to automate, such as downloading a file from a database, converting a format, or generating a standard plot; these the agent can take on with high autonomy. The second: tasks requiring supervision, such as a statistical analysis whose method you choose and whose execution the agent handles while you review the results. The third: tasks that are not automated, such as interpreting the significance of a biological result or making a strategic research decision. Success in deploying agents begins with placing each task in its correct category.

Calculating the Return on Investment

Not every automation is economically worthwhile. Building an agent for a task you perform once a year may consume more time than it saves. The practical rule: automate tasks that recur frequently

and take a reasonable amount of time each instance. A task you perform daily that takes ten minutes is more worth automating than one you perform monthly that takes an hour. As your tools and templates mature, the cost of building a new agent drops, widening the scope of what is worth automating. Start with high-frequency tasks and accumulate your experience and templates gradually.

Chapter 3

Setting Up Your Agentic Environment

Before we unleash agents on our tasks, we need a prepared environment. Setting up this environment in the context of bioinformatics has special considerations, the most important being the balance between capability and safety, especially when dealing with sensitive data. In this chapter we will survey the basic components of an effective and safe agentic work environment.

Choosing the Right Model for Scientific Tasks

Models differ in their capabilities, cost, and suitability for scientific work. For tasks requiring complex reasoning, such as interpreting a result or designing an analysis, you need a powerful model. For routine, repetitive tasks, such as formatting a file or extracting a field from text, a smaller, faster, cheaper model may suffice. It is wise to build a system that routes each task to the model most suitable for it, balancing quality and cost. Note that general models

may not know the latest specialized tools and terminology, so you may need to supply them with up-to-date references through retrieval.

Essential Tools for the Bioinformatician

The bioinformatics agent needs tools that connect it to the world of biological data. Among the most important: a tool to execute code (Python and R) in an isolated environment, a tool to access the APIs of biological databases such as NCBI services, a tool to search the scientific literature, and a tool to read files in their diverse biological formats (FASTA, FASTQ, VCF, BAM, and others). Each tool must be designed with a precise description the agent understands, and with clear bounds that prevent it from exceeding its privileges. Remember that the quality of your tools determines the ceiling of what your agent can accomplish.

Isolation and Safety: The Principle of Least Privilege

In a scientific environment that may handle precious or sensitive data, isolation becomes a necessity, not a luxury. Run the agent's

code in an isolated environment (a container or virtual environment) that can access only what it actually needs. Apply the principle of least privilege: do not grant the agent permission to delete files if it only needs to read them, nor internet access if its task is purely local. Make irreversible operations, such as deleting data or overwriting original files, require your explicit approval. These barriers are not obstacles to productivity, but safeguards that protect you from a costly mistake.

Managing Keys and Credentials

Many databases and services require access keys. Never place these keys directly in code, nor share them with the agent in the conversation text; instead store them in secure environment variables or dedicated secret vaults. Be especially careful when the agent handles health data subject to laws such as the U.S. HIPAA, since sending patient data to an external service may be a legal violation. In such cases, prefer models that run in your local environment or in an infrastructure compliant with privacy requirements. We will detail this aspect in the ethics chapter.

Chapter 4

Literature Review and Knowledge Synthesis Agents

Every research journey begins with understanding what preceded it. But the number of scientific papers in the life sciences grows at a rate exceeding any researcher's capacity to keep up: hundreds of thousands of papers annually in databases like PubMed. Here the literature-review agent becomes an indispensable tool, helping you scan quickly, identify the most important, and synthesize an initial picture. But it is a tool that must be used with great caution, as we shall see.

Intelligent Search and Filtering

The skilled agent does not merely pass your question to a search engine; it formulates multiple queries from different angles, searches specialized databases like PubMed through their APIs, then filters the results by relevance, recency, and importance, removes duplicates, and classifies what remains. Imagine asking: "Gather the

latest papers on the role of a particular gene in a particular disease over the last two years, and classify them by study type." The agent executes this in a few minutes, saving you hours of manual searching and giving you an organized starting point.

Summarization and Knowledge Synthesis

After gathering comes synthesis. The agent can read the abstracts of dozens of papers and extract from them the common patterns, the conflicting findings, and the research gaps. It can build a table comparing the methodologies and results of studies, or draw a conceptual map of the field. This "synthesis" is where the real value lies, transforming a heap of scattered papers into an organized understanding to build upon. But beware: the summary the agent provides is a starting point for your own reading, not a substitute for it.

The Danger of Hallucination in the Scientific Context

The most dangerous threat to the literature agent is "hallucination": fabricating a reference that does not exist, attributing a finding to a

study that did not state it, or mixing the results of different studies. This is a catastrophic danger in science, for citing a fictitious reference can destroy the credibility of an entire work. Prevention rests on a strict principle: do not trust any scientific claim the agent provides without verifying its original source yourself. Use retrieval techniques that bind the agent to base its statements on real, retrieved texts, always ask it to cite the precise source of each claim, then open the source and confirm. The sacred rule: verify, then verify.

Chapter 5

Data Retrieval and Database Querying

Biological data is distributed across an enormous network of public databases, each with its own interface, format, and method of querying. This diversity is both a blessing and a curse: a blessing because knowledge is available to all, and a curse because data retrieval requires mastering many tools. The skilled agent eases this burden by becoming an intelligent intermediary between you and these databases.

The Essential Databases

Every bioinformatician needs familiarity with the essential databases. The NCBI center holds treasures such as GenBank for nucleic acid sequences, PubMed for literature, and GEO and SRA for gene-expression and sequencing data. The Ensembl database provides rich genomic annotation. UniProt is a reference for proteins. PDB is for three-dimensional protein structures. Each of these databases has an

API that allows automated access to its content. The agent, when we equip it with tools that talk to these interfaces, can fetch what you need with a command in natural language instead of recalling the details of each interface.

From Natural Language to Structured Query

The most powerful thing the agent offers here is the translation from your intent into a technical query. You say, "I want all kidney cancer samples in GEO that used RNA sequencing and were published after a certain year," and the agent converts this request into a structured query with the correct fields and filters, executes it, and returns the results sorted. This conversion from intent to execution removes a technical barrier that long hindered researchers not specialized in programming, and puts the wealth of data within every researcher's reach.

Verifying the Integrity of Retrieved Data

Retrieved data is not always clean or complete. There may be samples with missing metadata, conflicting identifiers, or unexpected formats. A good agent does not merely fetch; it inspects what it

fetched: it checks the completeness of fields, reports any anomalies, and describes the state of the data to you before you build upon it. It is wise to always ask the agent for a brief report on the retrieved data: how many samples, how complete they are, and whether there is anything warranting attention. An analysis built on flawed data produces flawed results no matter how precise the processing.

Chapter 6

Sequence Analysis Agents

Sequence analysis is the beating heart of bioinformatics. From comparing a genetic sequence against a massive database, to aligning several sequences to reveal conserved regions, to predicting a protein's structure, these tasks recur daily in the world's laboratories. Many of them involve running specialized tools with precise parameters and interpreting their outputs, a pattern ideal for agent assistance.

Automating Similarity Search

Among the most well-known sequence-analysis tasks is similarity search using tools like BLAST, which compare your sequence against millions of known sequences to find the closest. This task involves repetitive steps: preparing the sequence, choosing the appropriate database, tuning sensitivity parameters, running the search, then interpreting the results. The agent can take on this entire chain: you provide a sequence and say "find the closest

matches and explain what they mean," and it prepares, runs, and interprets, highlighting the significant matches along with statistical confidence values.

Alignment and Conserved-Region Analysis

Multiple sequence alignment reveals the regions conserved across evolution, which are often regions of vital function. Running alignment tools, interpreting their results, and turning them into illustrative figures are tasks the agent masters. You can ask it to align a set of sequences, then mark the conserved positions, then generate a figure that highlights them visually. Thus you move from raw data to a biological insight in a few steps, while the final interpretation, why this region was conserved and what its functional significance is, remains in your hands as the researcher.

The Agent's Limits in Specialized Analysis

Despite agents' capability, there are limits that must be recognized. Specialized analysis may require deep expertise in choosing the right tool and the appropriate parameters for a particular data type. The agent may choose default parameters unsuitable for your case, or

apply a tool designed for a different context. You must therefore know what the agent is doing and why, rather than handing it the reins blindly. Use the agent to accelerate execution, but reserve the methodological decision for yourself, and review the parameters it chose before relying on the results.

Chapter 7

Pipeline Orchestration and Workflow Automation

An analysis in bioinformatics is rarely completed with a single tool. The reality is that most analyses are a chain of steps, each taking the output of its predecessor as its input: from quality control, to alignment, to counting, to statistical analysis. This chain is called a "pipeline," and managing it efficiently and reproducibly is among the hardest and most rewarding daily challenges when mastered.

Workflow Management Systems

The bioinformatics community has developed specialized systems for managing pipelines, the most well-known being Nextflow and Snakemake. These systems allow describing an analysis as a chain of interconnected rules, then automatically manage execution, parallelize independent steps, resume from where they stopped upon failure, and ensure reproducibility. But writing and maintaining these scripts requires expertise. Here the agent stands out: it can

help you write a new rule, fix an error in the workflow, or convert a scattered manual analysis into an organized, repeatable pipeline.

The Agent as Workflow Orchestrator

Beyond merely writing scripts, the agent can be a live orchestrator of the workflow. It monitors the pipeline's execution, detects when a step fails, reads the error message, suggests a fix, and may even execute the fix and resume after your approval. This pattern of "intelligent supervision" frees the researcher from watching long analyses hour by hour, and lets them intervene only when a human decision is needed. But always ensure the final word remains yours in any change that affects the analysis methodology or its results.

Reproducibility as a First Principle

The greatest thing the agent can offer a workflow is not speed, but reproducibility. Science that cannot be repeated is not science. Always ask your agent to document every step: the versions of the tools used, the specific parameters, and the source of each piece of data. Make it build self-contained pipelines that you or others can run later to yield the same results. A disciplined agent that makes

documentation and reproducibility an ingrained habit protects your research from the greatest blight of computational science: results that no one, not even their author, can reproduce.

Chapter 8

Data Wrangling and Statistical Analysis

Many bioinformaticians say that eighty percent of their time is lost in cleaning and preparing data, and only twenty percent in actual analysis. This "data wrangling", from unifying formats, to handling missing values, to merging tables, is tedious, repetitive, and error-prone, and it is exactly the burden agents excel at easing.

Cleaning and Preparing Data

The agent can inspect a dataset and describe its state to you: how many rows and columns, where the missing values are, and what the anomalous patterns are. It then writes the code, in Python or R, to clean it: unifying sample names, correcting types, handling missing values in the manner you choose, and merging multiple sources into a single consistent table. Most importantly, the agent makes this processing written in reviewable, re-runnable code, rather than

hidden manual edits in a spreadsheet that are impossible to trace or repeat.

Statistical Analysis Under Supervision

In statistical analysis the role of human supervision deepens. The agent can execute complex analyses, such as differential gene-expression analysis, clustering, or principal component analysis, and generate the code and outputs. But choosing the correct statistical method, ensuring its assumptions hold, and interpreting the significance of the results are all scientific decisions you must make. A common danger is that the agent applies a statistical test whose conditions are not met on your data, yielding a result that looks sound but is misleading. So: let the agent perform the computation, but you undertake the choice of method and the judgment of its validity.

Single-Cell Data Analysis as an Example

Take single-cell sequencing data analysis as an example of modern complexity. It involves many steps: quality control, normalization, dimensionality reduction, clustering, and cell-type identification.

Each step has choices and parameters that affect the outcome. The agent can build the entire analysis pipeline using standard libraries, generate figures at each stage, and explain what is happening. But it may err in identifying cell types or in choosing a threshold. So benefit from its speed of execution, and review every biological decision with your expert eye, for masterful analysis is a partnership between the machine's speed and the researcher's insight.

Chapter 9

Code Generation, Debugging, and Reproducibility

Programming is the daily currency of bioinformatics. Every analysis requires code: scripts to process data, functions for computations, and commands to run tools. Coding agents have caused a revolution here, becoming capable not merely of suggesting lines of code, but of writing entire modules, fixing intricate bugs, and refactoring existing code. This chapter is about how to benefit from this capability safely and masterfully.

From Description to Working Code

You can describe to the agent what you want in natural language, "write a function that reads a VCF file, filters variants by quality, and returns a table of the remaining variants", and it writes the complete code, explains it, and suggests tests for it. More powerfully, modern coding agents work in a loop: they write the code, run it, read the error if any, fix it, and repeat until it succeeds. This ability to

autonomously close the "write-run-fix" loop is what distinguishes the agent from a mere code generator, making it a true programming partner.

The Danger of Code That Runs but Is Wrong

The most dangerous thing about generated code is not the error that halts execution, but the silent error: code that runs without crashing, yet computes the wrong thing. The agent may confuse two definitions of some metric, or apply a transformation in the wrong place, producing numbers that look plausible but are mistaken. In science this is a catastrophe, because your results will be built on an erroneous computation without your knowing. Protection comes from testing the code on cases whose results you know in advance, and from reviewing the logic rather than merely confirming it "works." Do not trust that code is correct simply because it produced a number; make sure it produced the correct number.

Isolated Environments and Versioning

To ensure your code works today as it will a year from now, take care to isolate its environments and pin its versions. Ask the agent

to precisely specify the versions of every library you use, and to build an isolated environment that can be rebuilt later. Place your code under a version-control system, and have the agent help you write clear change messages. These habits seem formal but are the essence of reproducible computational science. An agent that helps you with discipline here offers your research a greater service than any momentary speedup.

Chapter 10

Visualization and Reporting Agents

A result that is not well presented is a wasted result. In bioinformatics, where the numbers are many and the relationships complex, good visualization is the bridge between analysis and understanding, and between the researcher and their audience. Preparing figures and reports, despite its importance, consumes much time in adjustment and formatting. Here agents offer great assistance.

Generating Scientific Figures

The agent can convert a natural-language description into a polished scientific figure. You say, "draw a heatmap of these genes' expression across the samples, with hierarchical clustering of rows and columns," and it writes the code, generates the figure, and adjusts colors, scales, and titles. You can request successive modifications in natural language: "enlarge the font," "change the color gradient," "add

a separator between groups." This interactive dialogue eliminates the tedious trial-and-error cycles of adjusting figures, and lets you reach the desired form quickly.

Reproducible Reports

The finest aspect of report preparation is for it to be reproducible: documents that combine text, code, and outputs, such that the entire report is regenerated when the data changes. Tools like R Markdown, Jupyter, and Quarto enable this, and the agent excels at building them. You can request a report that analyzes your data and generates the figures, tables, and interpretation in a single integrated document. When new data arrives, the report is regenerated at the press of a button. This pattern combines efficiency with reproducibility, and is among the most useful things the agent offers the researcher.

The Limits of Automated Interpretation

The agent may generate, alongside figures, textual interpretations of the results. Here caution is required. The agent may describe what it sees in the figure with a superficially correct description, but it may

leap to biological conclusions unsupported by evidence, or impute to a statistical correlation a causal meaning it has no right to. Deep scientific interpretation, linking the result to the biological context, estimating its significance, and distinguishing correlation from causation, remains the researcher's responsibility. Use the agent's description of figures as a starting point, but write the scientific interpretation yourself, for it is the essence of your contribution, not to be delegated.

Chapter 11

Multi-Agent Systems for Complex Analyses

When an analysis grows complex and ramified, a single agent trying to do everything may not suffice. Imagine a project requiring literature review, data retrieval, statistical analysis, visualization, and report writing. Distributing these roles among specialized agents that consult and complement one another may be more effective than loading a single agent with this entire burden. This is the logic of multi-agent systems in the bioinformatics context.

Teams of Specialized Agents

A team of agents mimicking a human research team can be designed. An agent specialized in research gathers literature and data. An agent specialized in analysis writes code and performs computations. An agent specialized in review inspects the others' outputs for errors. A coordinating agent distributes tasks and gathers results. This specialization allows each agent to carry in its

context only what concerns its task, so it is not crowded with all the project's details, and performs its role with greater mastery. Most importantly, the review agent adds a layer of verification that catches errors which might escape a solitary agent.

Mutual Verification as a Safety Layer

Among the most precious benefits of multiple agents in science is the mechanism of mutual verification. When one agent reviews another's work, oversight arises that catches errors early. For example, one agent might write a statistical analysis, then another inspects it, asking: is the method appropriate? Are the assumptions met? Is the interpretation supported? This automated questioning, though it does not dispense with the researcher's review, adds an extra line of defense against the silent errors that threaten the accuracy of computational science.

When to Avoid Complexity?

Despite the appeal of the idea, the same warning recurs: do not resort to multiple agents unless a simpler structure fails to solve the problem. Multi-agent systems are more complex to build and

maintain, costlier, and harder to debug. Many tasks that seem to need a full team are solved by a single well-designed agent with appropriate tools. The rule in bioinformatics as elsewhere: start simple, and add complexity only when it is proven that simplicity does not suffice. Complexity is a price paid when necessary, not an end sought for its own sake.

Chapter 12

Validation, Safety, and Scientific Rigor

This chapter is the conscience of the book. All the preceding capabilities become a danger if not disciplined by scientific rigor. The noble verse lays down the golden rule: if news comes to you from a source that may not be trustworthy, verify before you build upon it. The agent, however capable, is a source whose outputs must be verified, not an authority to be accepted without scrutiny.

Hallucination in the Life Sciences

Hallucination in bioinformatics takes malicious forms. The agent may invent the name of a gene or protein that does not exist, attribute a wrong function to a real gene, cite a fictitious paper, or claim an association between a disease and a factor without support. Its danger is that it comes with complete confidence and convincing scientific phrasing. There is no magic solution that eliminates hallucination entirely, but mitigating it is possible by binding the

agent to rely on real, retrieved sources, by requesting precise citation, and above all by independent human verification of every substantive claim.

Methodical Verification of Outputs

Verification is not a feeling but an organized method. Each type of output has a verification approach that suits it. Code is tested on cases with known results. Statistical results have their assumptions reviewed and are redone by an independent method when in doubt. Scientific claims are traced back to their original sources. Numbers are examined for biological plausibility. Make verification a fixed step in your workflow, not an afterthought. The greater a result's importance and consequences, the greater the rigor of verification required. This rigor is not an obstacle to speed; it is what separates science from mere convincing generation of text.

Avoiding Bias in Analysis

There is a danger subtler than outright error, and it is bias. The researcher may incline, consciously or unconsciously, to accept the agent's outputs that confirm their hypothesis and neglect what

contradicts it. The agent itself may carry biases from its training data. Prevention comes from defining success criteria before you see the results, testing your hypotheses seeking to refute rather than confirm them, and sometimes asking the agent to present the counterargument. Scientific rigor requires that we be harshest on the ideas we love, not that we seek someone to endorse them.

Chapter 13

Case Study: An End-to-End Analysis with Agents

After all the theory and principles, let us bring them together in an integrated practical scenario. We will follow a hypothetical researcher who wants to analyze gene-expression data to compare diseased samples with healthy ones, enlisting agents at every step while keeping scientific judgment in hand. This journey embodies how the preceding chapters integrate into a single piece of work.

From Question to Data

The researcher begins with a biological question: which genes are differentially expressed between patients and healthy individuals? They task the literature agent with scanning what has been published about the disease and its molecular mechanisms, obtaining an organized background within minutes, which they review themselves and verify its sources. Then they ask the data agent to search the GEO database for a suitable dataset, and the agent

retrieves the candidates and describes the state of each, so the researcher chooses the most suitable set based on their judgment. Note how the agent performs the heavy work while the scientific decision remains with the researcher.

From Data to Results

After choosing the data, the researcher asks the analysis agent to build a reproducible pipeline: quality control, then normalization, then differential expression analysis by a statistical method the researcher determines themselves based on the nature of the data. The agent executes and generates the list of differentially expressed genes with significance values. The researcher reviews the parameters and statistical assumptions and verifies the plausibility of the results. Then they ask the visualization agent for illustrative figures: a heatmap, a volcano plot, and a principal component analysis, so the results take shape in images that reveal the patterns.

From Results to Knowledge

Raw results are not the goal; understanding them is. The researcher enlists the agent in functional enrichment analysis to learn the

biological pathways to which the changed genes belong, obtaining a list of candidate pathways. But the final interpretation, which of these pathways has true significance in the context of the disease, how it relates to the original hypothesis, and what is worth experimental follow-up, this is the researcher's contribution, not to be delegated. Here the masterful researcher diverges from one who settles for the agent's outputs: the former builds knowledge, the latter gathers numbers.

From Knowledge to Report

Finally, the researcher asks the agent to build a reproducible report that brings together the methodology, code, results, and figures in a single document, documenting every version and parameter. The researcher writes the interpretation and discussion section themselves, for it is the core of their contribution. The result is work completed in days instead of weeks, without sacrificing rigor, and reproducible by any other researcher. This is the ideal partnership: the agent multiplies productivity, and the researcher ensures quality and meaning. This is the essence of what we wanted to teach in this book.

Chapter 14

Ethics, Data Privacy, and the Future

Our journey concludes with what must be at the forefront of every researcher's awareness: ethics and privacy. Data in the life sciences is not mere abstract numbers; much of it touches the privacy of human beings who entrusted us with the most precious thing they own: their health information and their genetic code. Deploying agents on this data imposes on us a doubled responsibility.

Health Data Privacy

Health data is subject to strict laws such as HIPAA in the United States, which protects patients' private information. The fundamental danger is that sending private health data to an external AI service may be a legal and ethical violation. Therefore, when you deal with data that identifies patients, avoid sending it to external services, and prefer models that run in your local environment or in a privacy-compliant infrastructure, or de-identify

the data before any processing. The noble verse "and do not spy" establishes a timeless principle: that people's privacy is an inviolable sanctity, not to be breached, even in the name of science.

Scientific Responsibility and Transparency

When you use agents in your research, full scientific responsibility remains on your shoulders, not on the tool. You are the one who guarantees the validity of the results, and you are the one held accountable for them. It is scientific honesty to be transparent about your use of artificial intelligence: to document in your methodology where you used it and how you verified its outputs. This transparency is not an admission of deficiency, but a commitment to the scientific integrity that allows others to evaluate and review your work. Science, in its essence, is a collective enterprise built on trust and verifiability.

The Future of Agentic Bioinformatics

Where is this field heading? We expect agents more specialized in the life sciences, trained for a deeper understanding of biology and its tools. We expect closer integration between agents, databases, and

tools through unified protocols, making the weaving of a complex analysis easier. We may see agents that propose hypotheses and design virtual experiments. But the essence of science will remain human: the original question, the creative intuition, the ethical judgment, and the responsibility for knowledge. The agent will magnify our capabilities, but it will not replace the scientific curiosity that drives research.

My final call remains: that our nation be a maker of these tools, not merely a consumer of them. We have minds and talents that qualify us to contribute to shaping this future, if we learn, build, and master. This book is but a seed we hope will bear researchers who harness the latest tools in the service of science and humanity. And God is the granter of success.

Conclusion

We have reached the conclusion of a practical journey in employing intelligent agents in the bioinformatician's daily life. We moved from understanding why these tools suit the nature of our work, to setting up the environment, then literature review, data retrieval, sequence analysis, pipeline orchestration, code generation, and visualization, all the way to an integrated case study. If there is one takeaway, it is that an agent multiplies a researcher's productivity when used in its right place, and with a scientific caution that preserves the accuracy of results.

The greatest danger in deploying these tools is not their incapacity, but blind trust in them. An agent may write a script that runs yet carries a hidden logical error, summarize a scientific paper misleadingly, or suggest a biological interpretation unsupported by evidence. For this reason the researcher's critical eye remains the last line of defense, and reproducibility and independent verification remain the standard of true science.

I hope the reader emerges with a firm conviction: that the future belongs not to the researcher whom the machine replaces, but to the researcher who masters working with it. Whoever excels at directing agents, verifying their outputs, and integrating them into their workflow will accomplish in a day what once required a week, and will be freed for what demands the creativity and scientific intuition the machine does not possess.

I ask God to make this work beneficial to every researcher striving to serve science and humanity, to render it knowledge from which people profit, and to grant our nation success in becoming a maker of the tools of knowledge rather than a consumer of them. My success is only through God. And God knows best.

And God knows best.

نبذة عن المؤلف

الدكتور فضل محمد الأكوع باحثٌ في البيولوجيا الحاسوبية والمعلوماتية الحيوية بجامعة ميشيغان (قسم الطب الباطني، شعبة أمراض الكلى) في مدينة آن أربور بالولايات المتحدة الأمريكية. تتركَّز اهتماماتُه البحثية في تطبيق الذكاء الاصطناعي والنماذج اللغوية الكبيرة على البيانات الطبية والحيوية، وتطوير أدواتٍ بحثيةٍ ذكية تخدم المجتمع العلمي، ومنها أداتُه البحثية في تحليل بيانات الأيض. وهو مؤسِّسُ مكتبةٍ رقميةٍ ثنائية اللغة، ومنتِجٌ لعشرات الكتب في مجالات الذكاء الاصطناعي والتقنية والدراسات الإسلامية والتاريخ والصحة والتربية، ويسعى من خلال إنتاجه المعرفي إلى تجسير الفجوة بين المعرفة التقنية المتقدمة والقارئ العربي.

Dr. Fadhl Mohammed Alakwaa is a researcher in computational biology and bioinformatics at the University of Michigan (Department of Internal Medicine, Division of Nephrology) in Ann Arbor, USA. His research focuses on applying artificial intelligence and large language models to biomedical data, and on developing intelligent research tools that serve the scientific community, including his research tool for metabolomics data analysis. He is the

founder of a bilingual digital library and the author of dozens of books spanning artificial intelligence, technology, Islamic studies, history, health, and parenting, and through his knowledge production he seeks to bridge the gap between advanced technical knowledge and the Arabic reader.

مؤلفات المؤلف

1. وكلاء الذكاء الاصطناعي في المعلوماتية الحيوية (2026)
2. إتقان الذكاء الاصطناعي الوكيل (2026)
3. نهاية الكون في القرآن الكريم – تحليل علمي شامل (2026)
4. مستقبل المعلوماتية الحيوية (2026)
5. تصنيف الآيات في القرآن الكريم (2026)
6. تأثير النماذج اللغوية الكبيرة على البحث العلمي (2026)
7. كيف تعمل النماذج اللغوية الكبيرة (2026)
8. دليل المهاجر اليمني الجديد إلى أمريكا (2026)
9. وادي السيليكون (2026)
10. أشهر المتنزهات الوطنية (2026)
11. الزواج الناجح (2026)
12. الإعجاز العلمي في القرآن الكريم (2026)
13. فن التربية (2026)

تمَّ بحمد الله إنجازُ هذا الكتاب

مكتبة فضل الأكوع الرقمية — 2026

أُنتج هذا الكتاب بمساعدة نموذج الذكاء الاصطناعي Claude Opus 4.8 من شركة Anthropic.

This book was produced with the assistance of Claude Opus 4.8 (Anthropic).

عن الكتاب

دليلٌ عمليٌّ موجَّهٌ للباحث في المعلوماتية الحيوية، يبيّن كيف يوظِّف الوكلاء الأذكياء في مهامه اليومية: من مراجعة الأدبيات واسترجاع البيانات، إلى تحليل التسلسلات وتنسيق خطوط الأنابيب وتوليد الشيفرة والتصوير، مع تأكيدٍ متواصلٍ على التحقق العلمي وخصوصية البيانات وقابلية إعادة الإنتاج.

عن المؤلف

د. فضل محمد الأكوع

باحثٌ في البيولوجيا الحاسوبية والمعلوماتية الحيوية بجامعة ميشيغان، شعبة أمراض الكلى، تتركَّز أبحاثُه في تطبيق الذكاء الاصطناعي على البيانات الطبية والحيوية.

أُنتج هذا الكتاب بمساعدة نموذج الذكاء الاصطناعي Claude Opus 4.8 من شركة Anthropic.
This book was produced with the assistance of Claude Opus 4.8 (Anthropic).

جميع الحقوق محفوظة للمؤلف © 2026